

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2005, 09:16:13 ; Search time 6.59898 Seconds
(without alignments)
565.610 Million cell updates/sec

Title: US-09-891-064A-2_COPY_89_138

Perfect score: 293

Sequence: 1 WDRGYGTSLGGSVGYPGG.....GYGYGYGYGYGYTDPRAAK 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCUTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	293	100.0	522	3	US-09-142-732-2
2	293	100.0	522	4	US-08-945-826-2
3	293	100.0	522	4	US-09-197-503-2
4	262.5	89.6	521	4	US-08-945-826-4
5	262.5	89.6	521	4	US-09-197-503-4
6	214	73.0	521	4	US-08-945-826-6
7	214	73.0	521	4	US-09-197-503-6
8	141	48.1	24	3	US-09-142-732-4
9	128	43.7	24	3	US-09-142-732-3
10	110.5	37.7	374	4	US-09-248-796A-19967
11	105	35.8	341	2	US-08-538-711A-8
12	105	35.8	341	3	US-08-725-027-8
13	105	35.8	341	4	US-09-542-552-8
14	105	35.8	353	2	US-08-538-711A-7
15	105	35.8	353	3	US-08-725-027-7
16	105	35.8	353	4	US-09-542-552-7
17	105	35.8	353	4	US-09-538-092-989
18	105	35.8	371	4	US-09-538-092-884
19	105	35.8	410	4	US-09-949-016-10345
20	105	35.8	410	4	US-09-949-016-10346
21	104	35.5	406	4	US-09-949-016-10343
22	103.5	35.3	193	4	US-09-270-767-43200
23	102.5	35.0	273	4	US-09-910-430-34
24	96.5	32.9	162	3	US-09-575-574-4
25	96.5	32.9	593	4	US-09-538-092-919
26	96	32.8	247	4	US-09-248-796A-14517
27	93	31.7	161	4	US-09-270-767-42771

28	92.5	31.6	564	4	US-09-949-016-6628	Sequence 6628, Ap
29	92.5	31.6	569	4	US-09-949-016-11035	Sequence 11035, A
30	92.5	31.6	569	4	US-09-949-016-11036	Sequence 11036, A
31	90.5	30.9	148	4	US-09-513-999C-5827	Sequence 5827, Ap
32	90.5	30.9	187	4	US-09-680-175-4	Sequence 4, Appl
33	90	30.7	168	3	US-09-512-342-20	Sequence 20, Appl
34	90	30.7	326	4	US-09-270-767-43241	Sequence 43241, A
35	89	30.4	637	4	US-09-949-016-8152	Sequence 8152, Ap
36	89	30.4	643	4	US-09-538-092-844	Sequence 844, Ap
37	88	30.0	241	4	US-09-270-767-40578	Sequence 40578, A
38	88	30.0	241	4	US-09-270-767-55794	Sequence 55794, A
39	87	29.7	435	4	US-09-949-016-8415	Sequence 8415, Ap
40	87	29.7	435	4	US-09-949-016-8792	Sequence 8792, Ap
41	86.5	29.5	274	4	US-09-976-594-417	Sequence 417, Ap
42	86	29.4	254	4	US-09-431-887-20	Sequence 20, Appl
43	86	29.4	254	4	US-09-431-887-21	Sequence 21, Appl
44	86	29.4	254	4	US-09-431-887-22	Sequence 22, Appl
45	86	29.4	378	4	US-10-164-595-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-09-142-732-2
; Sequence 2, Application US/09142732
; Patent No. 6252045
; GENERAL INFORMATION:
; APPLICANT: James M. Anderson
; APPLICANT: Christina M. Van Itallie
; TITLE OF INVENTION: Human Occludin, Its Uses and
; TITLE OF INVENTION: Enhancement of Drug Absorption Using Occludin Inhibitors
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yale University Medical School
; ADDRESSEE: Section of Digestive Diseases
; ADDRESSEE: Department of Internal Medicine
; STREET: 333 Cedar Street, LCI 105
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: United States of America
; ZIP: 06520-8057
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" 1.44 Mb diskette
; COMPUTER: IBM PC
; OPERATING SYSTEM: MS DOS
; SOFTWARE: Word Processing
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/142.732
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/05809
; FILING DATE: March 14, 1997
; APPLICATION NUMBER: U.S. 60/013,625
; FILING DATE: March 15, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary M. Krinsky
; REGISTRATION NUMBER: 32423
; REFERENCE/DOCKET NUMBER: 1751-P0016B.PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 203-324-6155
; TELEFAX: 203-327-1096
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 522
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: polypeptide
; FRAGMENT TYPE: complete sequence
; FEATURE:

2

NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:

; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19967
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-19967

Query Match 37.7%; Score 110.5; DB 4; Length 374;
Best Local Similarity 56.2%; Pred. No. 0.00021;
Matches 27; Conservative 3; Mismatches 9; Indels 9; Gaps 5;

QY 4 GYGTSLLGSGVCP-YG-----GSGFGS--YGSY-GYGYGYGYYG 43
DB 96 GYGSSMYGGYGSSMYGNSMYGGYGSSMYGGYGRGSGYG-GYGSY 142

RESULT 11
US-08-538-711A-8
; Sequence 8, Application US/08538711A
; Patent No. 5994062
; GENERAL INFORMATION:
; APPLICANT: MULSHINE, JAMES, L.
; TITLE OF INVENTION: AN EPITHELIAL PROTEIN AND
; TITLE OF INVENTION: DNA THEREOF FOR USE IN EARLY CANCER DETECTION
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154

; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/538,711A
; FILING DATE: 02-OCT-1995

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: KATHRYN M. BROWN
; REGISTRATION NUMBER: 34,556
; REFERENCE/DOCKET NUMBER: 2026-4201

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 341
; TYPE: Amino Acid
; STRANDEDNESS: Unknown
; TOPOLOGY: Linear
; MOLECULE TYPE: peptide
US-08-538-711A-8

Query Match 35.8%; Score 105; DB 2; Length 341;

Best Local Similarity 50.0%; Pred. No. 0.00076;
Matches 23; Conservative 2; Mismatches 13; Indels 8; Gaps 2;
QY 4 GYGTSLLGSGVCPYVGSGFGSGYSGYGVY----GYGVGYGYTD 45
DB 249 GYG-----GGRGGYGGGPGYGNQGGYGGYDNYGGGNYGSGNYND 290

RESULT 12
US-08-725-027-8
; Sequence 8, Application US/08725027
; Patent No. 6251586
; GENERAL INFORMATION:
; APPLICANT: MULSHINE, JAMES, L.
; APPLICANT: TOCKMAN, MELVYN, S.
; TITLE OF INVENTION: AN EPITHELIAL PROTEIN AND
; TITLE OF INVENTION: DNA THEREOF FOR USE IN EARLY CANCER DETECTION
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154

; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/725,027
; FILING DATE: 02-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US08/538,711
; FILING DATE: 02-OCT-1995

; ATTORNEY/AGENT INFORMATION:
; NAME: KATHRYN M. BROWN
; REGISTRATION NUMBER: 34,556
; REFERENCE/DOCKET NUMBER: 2026-4201US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 341
; TYPE: Amino Acid
; STRANDEDNESS: Unknown
; TOPOLOGY: Linear
; MOLECULE TYPE: peptide
US-08-725-027-8

Query Match 35.8%; Score 105; DB 3; Length 341;
Best Local Similarity 50.0%; Pred. No. 0.00076;
Matches 23; Conservative 2; Mismatches 13; Indels 8; Gaps 2;
QY 4 GYGTSLLGSGVCPYVGSGFGSGYSGYGVY----GYGVGYGYTD 45
DB 249 GYG-----GGRGGYGGGPGYGNQGGYGGYDNYGGGNYGSGNYND 290

RESULT 13
US-09-542-552-8
; Sequence 8, Application US/09542552
; Patent No. 650625
; GENERAL INFORMATION:
; APPLICANT: MULSHINE, JAMES, L.
; TITLE OF INVENTION: AN EPITHELIAL PROTEIN AND
; TITLE OF INVENTION: DNA THEREOF FOR USE IN EARLY CANCER DETECTION
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE

;; CITY: NEW YORK
;; STATE: NEW YORK
;; COUNTRY: USA
;; ZIP: 10154
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: FLOPPY DISK
;; COMPUTER: IBM PC COMPATIBLE
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: ASCII
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/542,552
;; FILING DATE: 03-APR-2000
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/538,711
;; FILING DATE: 02-OCT-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: KATHRYN M. BROWN
;; REGISTRATION NUMBER: 34,556
;; REFERENCE/DOCKET NUMBER: 2026-4201
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 758-4800
;; TELEFAX: (212) 751-6849
;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 341
;; TYPE: Amino Acid
;; STRANDEDNESS: Unknown
;; TOPOLOGY: Linear
;; MOLECULE TYPE: peptide
US-09-542-552-8

Query Match 35.8%; Score 105; DB 4; Length 341;
Best Local Similarity 50.0%; Pred. No. 0.00076;
Matches 23; Conservative 2; Mismatches 13; Indels 8; Gaps 2;

Qy 4 GYGTSLGGSGVGPYGGSGFGSGYGYG---GYGYGGYTD 45
Db 249 GYG----GGRGGYGGGPGYGNQGGYGGYDNYGGNGSGNYND 290

RESULT 14
US-08-538-711A-7
; Sequence 7, Application US/08538711A
; Patent No. 5994062
; GENERAL INFORMATION:
; APPLICANT: MULSHINE, JAMES, L.
; TITLE OF INVENTION: AN EPITHELIAL PROTEIN AND
; DNA THEREOF FOR USE IN EARLY CANCER DETECTION
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/538,711A
; FILING DATE: 02-OCT-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: KATHRYN M. BROWN
; REGISTRATION NUMBER: 34,556
; REFERENCE/DOCKET NUMBER: 2026-4201

;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 758-4800
;; TELEFAX: (212) 751-6849
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 353
;; TYPE: Amino Acid
;; STRANDEDNESS: Unknown
;; TOPOLOGY: Linear
;; MOLECULE TYPE: peptide
US-08-538-711A-7

Query Match 35.8%; Score 105; DB 2; Length 353;
Best Local Similarity 50.0%; Pred. No. 0.00078;
Matches 23; Conservative 2; Mismatches 13; Indels 8; Gaps 2;

Qy 4 GYGTSLGGSGVGPYGGSGFGSGYGYG---GYGYGGYTD 45
Db 261 GYG----GGRGGYGGGPGYGNQGGYGGYDNYGGNGSGNYND 302

RESULT 15
US-08-725-027-7
; Sequence 7, Application US/08725027
; Patent No. 6251586
; GENERAL INFORMATION:
; APPLICANT: MULSHINE, JAMES, L.
; APPLICANT: TOCKMAN, MELVIN, S.
; TITLE OF INVENTION: AN EPITHELIAL PROTEIN AND
; DNA THEREOF FOR USE IN EARLY CANCER DETECTION
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/725,027
; FILING DATE: 02-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US08/538,711
; FILING DATE: 02-OCT-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KATHRYN M. BROWN
; REGISTRATION NUMBER: 34,556
; REFERENCE/DOCKET NUMBER: 2026-4201US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 353
; TYPE: Amino Acid
; STRANDEDNESS: Unknown
; TOPOLOGY: Linear
; MOLECULE TYPE: peptide
US-08-725-027-7

Query Match 35.8%; Score 105; DB 3; Length 353;
Best Local Similarity 50.0%; Pred. No. 0.00078;
Matches 23; Conservative 2; Mismatches 13; Indels 8; Gaps 2;

Qy 4 GYGTSLGGSGVGPYGGSGFGSGYGYG---GYGYGGYTD 45
Db 261 GYG----GGRGGYGGGPGYGNQGGYGGYDNYGGNGSGNYND 302

Search completed: April 1, 2005, 10:04:44
Job time : 6.59898 secs

BEST AVAILABLE COPY

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2005, 09:31:32 ; Search time 24.2809 Seconds
(without alignments)
1054.491 Million cell updates/sec

Title: US-09-891-064A-2_COPY_89_138

Perfect score: 293
Sequence: 1 WDRGYGTSLLGSGVGYGG.....GYGYGYGYGYGTYDPRAAK 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	293	100.0	522	1 OCLN_HUMAN	Q16625 homo sapien
2	262.5	89.6	521	1 OCLN_CANFA	Q28269 canis famil
3	262.5	89.6	560	2 O9N0W3	Q9N0W3 canis famil
4	238	81.2	449	2 Q6P84	Q6P84 bos taurus
5	233	79.5	523	2 Q6E6T5	Q6E6T5 rattus norv
6	233	79.5	523	2 Q9Z303	Q9Z303 rattus norv
7	214	73.0	521	1 OCLN_MOUSE	Q61146 mus musculu
8	128.5	43.9	195	1 ROAB_ARTSA	P80350 artemia aal
9	127.5	43.5	54	2 O08636	O08636 mus musculu
10	123.5	42.2	188	2 Q772P9	Q772P9 brachydanio
11	120.5	41.1	62	2 O08632	O08632 mus musculu
12	120.5	41.1	141	2 Q92510	Q92510 mus musculu
13	119	40.6	493	1 OCLN_XENLA	Q9PUN1 xenopus lae
14	118.5	40.4	79	1 KRHA_RABIT	Q02957 oryctolagus
15	118	40.3	88	1 K191_HUMAN	Q81UB9 homo sapien
16	116.5	39.8	81	2 Q724W3	Q724W3 homo sapien
17	116	39.6	180	2 Q81J1	Q81J1 fugu rubrip
18	115.5	39.4	209	2 Q9VX73	Q9VX73 drosophila
19	115.5	39.4	236	2 Q6NP84	Q6NP84 drosophila
20	115.5	39.4	497	2 Q6NRR0	Q6NRR0 xenopus lae
21	115	39.2	132	2 Q701T7	Q701T7 anopheles g
22	114	38.9	86	2 Q81I16	Q81I16 mus musculu
23	114	38.9	121	2 Q703F5	Q703F5 anopheles g
24	113.5	38.7	184	2 Q701T9	Q701T9 anopheles g
25	112.5	38.4	77	2 Q08631	Q08631 mus musculu
26	112.5	38.4	78	2 O09048	O09048 mus musculu
27	112.5	38.4	78	2 Q925H3	Q925H3 mus musculu
28	112.5	38.4	381	2 Q9GPD9	Q9GPD9 ixodes ricin
29	112.5	38.4	500	2 Q6NXX9	Q6NXX9 brachydanio
30	112	38.2	210	2 Q9AAZ0	Q9AAZ0 caulobacter
31	112	38.2	404	2 Q22791	Q22791 arabidopsis

32	111	37.9	440	2	Q9V6M1	Q9V6M1 drosophila
33	111	37.9	541	2	Q87BZ7	Q87BZ7 xylella fas
34	110.5	37.7	87	2	Q925H2	Q925H2 mus musculu
35	110.5	37.7	87	2	Q925H6	Q925H6 mus musculu
36	110.5	37.7	102	2	Q9176	Q9176 caenorhabdi
37	110.5	37.7	127	1	SGP1_CHRVI	OS02057 chromatiu
38	110.5	37.7	1790	2	Q81816	Q81816 aplysia cal
39	110	37.5	84	2	Q925H7	Q925H7 mus musculu
40	110	37.5	252	2	Q7VEJ9	Q7VEJ9 prochloroco
41	109.5	37.4	240	2	Q6B8G7	Q6B8G7 ixodes paci
42	109	37.2	119	2	Q922U9	Q922U9 ixodes scap
43	109	37.2	244	2	Q75A59	Q75A59 ashbya goss
44	109	37.2	324	2	Q7ZY9	Q7ZY9 xenopus lae
45	109	37.2	504	1	OCLN_CHICK	Q91049 gallus gall

ALIGNMENTS

RESULT 1
OCLN_HUMAN STANDARD; PRT; 522 AA.
AC Q16625; Q8N6K1;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Occludin.
GN Name=OCLN;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon carcinoma;
RX MEDLINE=96181088; PubMed=8601611; DOI=10.1083/jcb.133.1.43;
RA Ando-Akatsuka Y., Saitou M., Hirase T., Kishi M., Sakakibara A.,
RA Itoh M., Yonemura S., Furuse M., Tsukita S.;
RT "Interspecies diversity of the occludin sequence: cDNA cloning of
RT human, mouse, dog, and rat-kangaroo homologues.";
RL J. Cell Biol. 133:43-47(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA van Itallie C.M., Fanning A.S., Anderson J.M.;
RL Submitted (MAY-1996) to the ENBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Fukasawa M., Toyota T., Yoshitsugu K., Yoshikawa T.;
RL "Genomic structure of occludin gene";
RN Submitted (JUL-2001) to the ENBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, and Lung;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.M., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzly D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Smith J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human

DR InterPro; IPR008253; Marvel.
DR InterPro; IPR002958; Occludin.
DR InterPro; IPR010844; Occludin_ELL.
DR Pfam; PF01284; MARVEL; 1.
DR Pfam; PF07303; Occludin_ELL; 1.
DR PRINTS; PR01258; OCCLUDIN.
FT NON_TER 1
SQ SEQUENCE 449 AA; 50627 MW; 235BE8259CF05D1C CRC64;

Query Match 81.2%; Score 238; DB 2; Length 449;
Best Local Similarity 84.3%; Pred. No. 5.2e-15;
Matches 43; Conservative 1; Mismatches 5; Indels 2; Gaps 2;

Qy 1 WDRGYGTSILLGSGVGYPGSGRGSFGSYGYGYGYGYGVG-GYTPPRAAK 50
|||||||:| | || || || || || || || || || || || || || || || || ||
Dd 16 WDRGYGT-LMGAGWNPYAGSAFSGYSYGSGYGYGYGYGTGTDPPRAAK 65

RESULT 5

ID	Q6P6T5	PRELIMINARY;	PRT;	523 AA.
AC	O6P6T5;			
DT	05-JUL-2004	(TrEMBLrel. 27, Created)		
DT	05-JUL-2004	(TrEMBLrel. 27, Last sequence update)		
DT	05-JUL-2004	(TrEMBLrel. 27, Last annotation update)		
DE	Occludin.			
GN	Names=Occln;			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=101116;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RC	TISSUE=Prostate;			
RX	MEDLINE-22388257; PubMed-12477932; DOI=10.1073/pnas.242603899;			
RA	Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,			
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,			
RA	Brownstein M.J., Udell T.B., Toshnyuki S., Carninci P., Prange C.,			
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,			
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,			
RA	Villalon D.K., Murny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Fahay J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,			
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,			
RA	Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,			
RA	Jones S.J., Marra M.A.;			
RT	"generation and initial analysis of more than 15,000 full-length human			
RT	and mouse cDNA sequences."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Prostate;			
RC	Straubeberg R.;			
RL	Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.			
RL	EMBL; BC062037; AAH62037.1; -.			
DR	GO; GO:0016021; C:integral to membrane; IEA.			
DR	GO; GO:0005923; C:right junction; IEA.			
DR	GO; GO:0005198; F:structural molecule activity; IEA.			
DR	InterPro; IPR008253; Marvel.			
DR	InterPro; IPR002958; Occludin.			
DR	InterPro; IPR010844; Occludin_ELL.			
DR	Pfam; PF01284; MARVEL; 1.			
DR	Pfam; PF07303; Occludin_ELL; 1.			
DR	PRINTS; PR01258; OCCLUDIN.			
SQ	SEQUENCE 523 AA; 59186 MW; A31B00758AEASB5 CRC64;			

Query Match 79.5%; Score 233; DB 2; Length 523;

Best Local Similarity 80.4%; Pred. No. 1.8e-14;
Matches 41; Conservative 3; Mismatches 5; Indels 2; Gaps 2;

QY 1 WDRGYGTLLGSGVGYPGSGFGSGYSGY-GYGYGYGYGYGYTDPRAAK 50
DB 89 WDRAYGTGTFGGSMNYPY-GSGFGSGYGGFGGYGYGYGYGYTDPRAAK 138

RESULT 6

QZ303 ID Q92303 PRELIMINARY; PRT; 523 AA.
AC Q92303;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Occludin.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]

SEQUENCE FROM N.A.
RC STRAIN=SD; TISSUE=Liver;
RA Kokai Y., Kuwahara K., Atsumi S., Mori M.;

RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016425; BAA36681.1; --

DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005923; C:tight junction; IEA.

DR InterPro; IPR008253; Marvel.
DR InterPro; IPR002958; Occludin_ELL.

DR Pfam; PF01284; MARVEL; 1.
DR Pfam; PF07303; Occludin_ELL; 1.

DR PRINTS; PR01258; OCCLUDIN.
SQ SEQUENCE 523 AA; 59214 MW; 5E1E007568AE5B1 CRC64;

Query Match 79.5%; Score 233; DB 2; Length 523;
Best Local Similarity 80.4%; Pred. No. 1.8e-14;

Matches 41; Conservative 3; Mismatches 5; Indels 2; Gaps 2;

QY 1 WDRGYGTLLGSGVGYPGSGFGSGYSGY-GYGYGYGYGYGYTDPRAAK 50
DB 89 WDRAYGTGTFGGSMNYPY-GSGFGSGYGGFGGYGYGYGYGYTDPRAAK 138

RESULT 7

OCNL MOUSE
ID OCLN MOUSE STANDARD; PRT; 521 AA.
AC Q61146;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Occludin.
GN Name=Occln; Synonyms=Ocl;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Ando-Akatsuka Y., Saitou M., Hirase T., Kishi M., Sakakibara A.,
RA Itoh M., Yonemura S., Furuse M., Tsukita S.;

RT "Interspecies diversity of the occludin sequence: cDNA cloning of
human, mouse, dog, and rat-kangaroo homologues.";

RL J. Cell Biol. 133:43-47(1996).
RN [2]

SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Ovary, and Uterus;

RA MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RX Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bodo H., Kondo S.,

RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matusda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sadelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
CC -!- FUNCTION: May play a role in the formation and regulation of the
tight junction (TJ) paracellular permeability barrier.
CC -!- SUBUNIT: Interacts with TJPI/ZO1 and with VAPA (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: Localized at tight junctions of both
epithelial and endothelial cells. Highly expressed in the testis,
kidney, lung, liver and brain. Not detected in skeletal muscle,
spleen and heart.
CC -!- DEVELOPMENTAL STAGE: Found diffusely on the lateral membranes of
Sertoli cells in the early prepubertal period. With development, it
became gradually concentrated at the most basal regions of Sertoli
cells.
CC -!- DOMAIN: The C-terminal is cytoplasmic and is important for
interaction with ZO-1. Necessary for the tight junction
localization. Involved in the regulation of the permeability
barrier function of the tight junction (By similarity).
CC -!- PTM: Phosphorylated (By similarity).
CC -!- SIMILARITY: Belongs to the ELL / occludin family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.ebi.ac.uk/ebis-sib.ch/announcement/>
or send an email to license@sib-sib.ch).

EMBL; U49185; AAC52515.1; --
DR EMBL; AK019880; BAB31900.1; --
DR MGD; MGI:106183; Occln.
DR GO; GO:0005886; C:plasma membrane; IEA.
DR GO; GO:0005923; C:tight junction; IEA.
DR GO; GO:0005515; F:protein binding; IPI.
DR InterPro; IPR008253; Marvel.
DR InterPro; IPR002958; Occludin.
DR InterPro; IPR010844; Occludin_ELL.
DR Pfam; PF07303; Occludin_ELL; 1.
DR PRINTS; PR01258; OCCLUDIN.
DR PRINTS; PR01258; OCCLUDIN.
KW Coiled coil; Phosphorylation; Tight junction; Transmembrane.
DR DOMAIN 1 66 Cytoplasmic (Potential).
DR TRANSMEM 67 89 Potential.
DR DOMAIN 90 133 Extracellular (Potential).
DR TRANSMEM 134 158 Potential.
DR DOMAIN 159 168 Cytoplasmic (Potential).
FT


```
FT TRANSMEM 169 193 Potential.
FT DOMAIN 194 241 Extracellular (Potential).
FT TRANSMEM 242 263 Potential.
FT DOMAIN 264 521 Cytoplasmic (Potential).
FT DOMAIN 92 129 Gly/Tyr-rich.
FT DOMAIN 307 310 Poly-Pro.
FT DOMAIN 424 488 Coiled coil (Potential).
SQ SEQUENCE 521 AA; 59000 MW; 21D62F3708B33E7D CRC64;

Query Match
Best Local Similarity 73.0%; Score 214; DB 1; Length 521;
Matches 41; Conservative 1; Mismatches 5; Indels 4; Gaps 3;

QY 1 WDRGYGTSLLGSGVGYPGSGFGSGYSGYGYG-YGYGYGYTDPRAAK 50
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 89 WDRGYGTGLFGSLNYPY--SGFG-YGGYGGYGGYGYGYGYTDPRAAK 136

RESULT 8
ROAB ARTSA
ID ROAB ARTSA STANDARD; PRT; 195 AA.
AC P80350;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Heterogeneous nuclear ribonucleoprotein A/B (P38) (Fragments).
OS Artemia salina (Brine shrimp).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Branchiopoda; Anostraca;
OC Artemiidae; Artemia.
OX NCBI_TaxID=85549;
RN [1]
RP SEQUENCE.
RX MEDLINE=95081083; PubMed=7527388;
RA Pyge S., Slegers H., Moens L., Merlevede W., Goris J.;
RT "Tyrosine phosphorylation of a M(r) 38,000 A/B-type hnRNP protein
    selectively modulates its RNA binding.";
RL J. Biol. Chem. 269:31457-31465(1994).
CC -!- FUNCTION: May regulate mRNA translation and stability. It binds to
    poly(A) and poly(U) regions of RNA. This binding is inhibited when
    the protein is phosphorylated.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear. At the onset of the
    cryptoblastic phase, when transcription is blocked, it accumulates
    in the cytoplasm.
CC -!- PTM: Extensively phosphorylated on tyrosine residues.
CC -!- SIMILARITY: Contains at least 1 RNA recognition motif (RRM)
    domain.
DR InterPro; IPR000504; RNA_rec mot.
DR PROSITE; PS50102; RRM; PARTIAL.
KW Direct protein sequencing; Methylation; Nuclear protein;
KW Phosphorylation; Ribonucleoprotein; RNA-binding.
FT NON_TER 1 1
FT NON_CONS 15 16 RNA-binding (RRM).
FT DOMAIN 32 >48
FT NON_CONS 48 49
FT NON_CONS 58 59
FT NON_CONS 73 74
FT NON_CONS 83 84
FT NON_CONS 96 97
FT DOMAIN 120 195
FT MOD_RES 119 119 Gly-rich.
FT MOD_RES 122 122 Asymmetric dimethylarginine.
FT MOD_RES 173 173 Asymmetric dimethylarginine.
FT MOD_RES 173 173 Phosphoserine (by CK2).
FT MOD_RES 192 192 Asymmetric dimethylarginine.
FT NON_CONS 125 126
FT NON_TER 135 135
SQ SEQUENCE 195 AA; 20604 MW; CA062DA8520B3DB CRC64;

Query Match
Best Local Similarity 43.9%; Score 128.5; DB 1; Length 195;
Matches 27; Conservative 3; Mismatches 11; Indels 3; Gaps 3;

QY 4 GYGTSLLGSGVGYPGSGFGSGYSGYGYG-YGYGYGYTDP 46
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 149 GYGDDAYGGA-GYDYGYSGYGGYGGY-EGYNGYGGYGGSGP 190

RESULT 9
O08636 PRELIMINARY; PRT; 54 AA.
ID O08636
AC O08636;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE High-glycine tyrosine keratin type II.3 (Fragment).
GN Name=Krtap6-3;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=98043760; PubMed=9374545; DOI=10.1074/jbc.272.48.30512;
RA Aoki N., Ito K., Ito M.;
RT "Isolation and characterization of mouse high-glycine/tyrosine
    proteins.";
RL J. Biol. Chem. 272:30512-30518(1997).
DR EMBL; D89901; BAA19687.1; -.
DR MGD; MGI:1330279; Krtap6-3.
DR GO; GO:0005882; C:Intermediate filament; IEA.
KW Keratin.
FT NON_TER 1 1
SQ SEQUENCE 54 AA; 5729 MW; 2E7880B8822C1787 CRC64;

Query Match
Best Local Similarity 43.5%; Score 127.5; DB 2; Length 54;
Matches 27; Conservative 2; Mismatches 6; Indels 7; Gaps 3;

QY 3 RGYGTSLLGSGVGYPGSGFGSGYSGYGYGYGYGYG-YGYGY 43
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2 RGY-----YGLGYGYGLGYG-YGSGYGGYGGYGGYGGY 37

RESULT 10
Q7T2P9 PRELIMINARY; PRT; 188 AA.
ID Q7T2P9
AC Q7T2P9;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Prion-like protein 1 precursor.
GN Name=PrPL-P1-like;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Hepatopancreas;
RA Strumbo B., Sangiorgio L., Bolis L.C., Ronchi S., Simonin T.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ490524; CAD35502.1; -.
KW Prion; Signal.
FT SIGNAL 1 24
FT CHAIN 25 161 prion-like protein 1.
SQ SEQUENCE 188 AA; 18700 MW; F1753B695FE39FC CRC64;

Query Match
Best Local Similarity 42.2%; Score 123.5; DB 2; Length 188;
Matches 21; Conservative 9; Mismatches 9; Indels 1; Gaps 1;

QY 4 GYGTSLLGSGVGYPGSGFGSGYSGYGYGYGYGYGYGY 43
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 106 GHGLSSM-GRFGYGYGGYGGYGGYGGYGGYGGHGGH 144
```

```

RESULT 11
O08632
ID O08632 PRELIMINARY; PRT; 62 AA.
AC O08632;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Glycine tyrosine-rich hair protein.
GN Name=Krtap8-2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=98043760; PubMed=9374545; DOI=10.1074/jbc.272.48.30512;
RA Aoki N., Ito K., Ito M.;
RT "Isolation and characterization of mouse high-glycine/tyrosine
RT proteins.";
RL J. Biol. Chem. 272:30512-30518 (1997).
DR EMBL; D86422; BAA19683.1; -.
DR MGD; MGI:1330295; Krtap8-2.
SQ SEQUENCE 62 AA; 6688 MW; 493A2EB2B0C219A5 CRC64;

Query Match 41.1%; Score 120.5; DB 2; Length 62;
Best Local Similarity 57.4%; Pred. No. 0.00016;
Matches 27; Conservative 4; Mismatches 9; Indels 7; Gaps 4;

QY 4 GYGTSLLG-GSYGYPYG-GSGFGSGYSGYGVG-YGYGVG-YYYGGY 43
DB 10 GLGSGIRFGNLGYGCGGFGYGSYGRYGYGPRPLYGGY 56

RESULT 12
Q92510
ID Q92510 PRELIMINARY; PRT; 141 AA.
AC Q92510;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Keratin-associated protein 16.1.
GN Name=Krtap16-1; Synonyms=Krtap16.1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB;
RX MEDLINE=21185977; PubMed=11290294;
RA Tkatchenko A.V., Visconti R.P., Shang L., Papenbrock T., Pruett N.D.,
RA Ito T., Ogawa M., Awgulewitsch A.;
RT "Overexpression of Hoxc13 in differentiating keratinocytes results in
RT downregulation of a novel hair keratin gene cluster and alopecia.";
RL Development 128:1547-1558 (2001).
DR EMBL; AF345291; AAK52889.1; -.
DR HSSP; P10968; 7WGA.
DR MGD; MGI:2157572; Krtap16-1.
SQ SEQUENCE 141 AA; 13738 MW; EF58E63AD0BEF3E0 CRC64;

Query Match 41.1%; Score 120.5; DB 2; Length 141;
Best Local Similarity 53.8%; Pred. No. 0.00033;
Matches 28; Conservative 3; Mismatches 8; Indels 13; Gaps 4;

QY 4 GYGTSLLG-GSYGVG-PYGGSGFGS-----YGSYGVGYGVGYGVGY 43
DB 63 GYGSGYGVGFGFGFYGSYGVGFGFGFGYGVGFGFGFGYGVGFGFGY 113

RESULT 13
OCLN_XENLA
ID Q9PUN1; STANDARD; PRT; 493 AA.
AC Q9PUN1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Occludin.
GN Name=OCLN;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A., AND PHOSPHORYLATION SITES THR-375 AND SER-379.
RC TISSUE=Ovary;
RX MEDLINE=99421641; PubMed=10491082;
RA Cordenonsi M., Turco F., D'Atri F., Hammar E., Martinucci G.,
RA Meggio F., Citi S.;
RT "Xenopus laevis occludin. Identification of in vitro phosphorylation
RT sites by protein kinase CK2 and association with cingulin.";
RL Eur. J. Biochem. 264:374-384 (1999).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=98034414; PubMed=9365283;
RA Cordenonsi M., Mazzon E., De Rigo L., Baraldo S., Meggio F., Citi S.;
RT "Occludin dephosphorylation in early development of Xenopus laevis.";
RL Cell Sci. 110:3131-3139 (1997).
CC -I- FUNCTION: Probably plays a role in the formation and regulation of
CC the tight junction (TJ) paracellular permeability barrier.
CC -I- SUBUNIT: Interacts in vitro with cingulin, possibly directly.
CC Interacts with ZO-1 (By similarity).
CC -I- SUBCELLULAR LOCATION: Integral membrane protein.
CC -I- TISSUE SPECIFICITY: Localized at tight junctions of both
CC epithelial and endothelial cells.
CC -I- DEVELOPMENTAL STAGE: A maternally synthesized protein. Found in
CC granules in the peripheral cytoplasm in the fertilized egg, it
CC localizes first to the basolateral membrane, then to tight
CC junctions after cingulin and ZO-1. Nascent tight junctions are in
CC place by the two-cell stage. The maternal form is more highly
CC phosphorylated than the form detected in later developmental
CC stages.
CC -I- DOMAIN: The C-terminus is cytoplasmic and is important for
CC interaction with ZO-1. Necessary for the tight junction
CC localization. Involved in the regulation of the permeability
CC barrier function of the tight junction (By similarity).
CC -I- SIMILARITY: Belongs to the ELL / occludin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)
CC -----
DR EMBL; AF170275; AAD53725.1; -.
DR IntAct; Q9PUN1; -.
DR InterPro; IPR008253; Marvel.
DR InterPro; IPR002958; Occludin.
DR InterPro; IPR010844; Occludin_ELL.
DR Pfam; PF01284; MARVEL; 1.
DR Pfam; PF07303; Occludin_ELL; 1.
DR PRINTS; PR01358; OCLLUDIN.
DR Coiled coil; Phosphorylation; Tight junction; Transmembrane.
FT DOMAIN 1 47 Cytoplasmic (Potential).
FT TRANSMEM 48 70 Potential.
FT DOMAIN 71 116 Extracellular (Potential).
FT TRANSMEM 117 141 Potential.
FT DOMAIN 142 151 Cytoplasmic (Potential).
FT TRANSMEM 152 176 Potential.
FT DOMAIN 177 224 Extracellular (Potential).
FT TRANSMEM 225 246 Potential.

```

```

FT DOMAIN      247 493      Cytoplasmic (Potential).
FT DOMAIN      396 428      Coiled coil (Potential).
FT MOD_RES     375 375      Phosphothreonine (by CK2) (in vitro).
FT MOD_RES     379 379      Phosphoserine (by CK2) (in vitro).
SQ SEQUENCE    493 AA; 55111 MW; 9694CD302BEBEDE CRC64;

Query Match    40.6%; Score 119; DB 1; Length 493;
Best Local Similarity 50.9%; Pred. No. 0.0014;
Matches 29; Conservative 4; Mismatches 10; Indels 14; Gaps 5;

QY 1 WDRGVTSLGSGVGPYGGSGFGSGY-CYGY-----GYGYGG-YTDPRAAK 50
Db 70 WD-----LDITGSMGY---GMGSGSYGGYTGFGGSGMGLGFAYGNYTDPRAAK 119

RESULT 14
KRHA_RABIT
ID KRHA_RABIT STANDARD; PRT; 79 AA.
AC Q02957;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Keratin, glycine/tyrosine-rich of hair.
GN Name=KAP6-1;
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93179467; PubMed=7680040;
RA Frattini A., Powell B.C., Rogers G.E.;
RT "Sequence, expression, and evolutionary conservation of a gene
RT encoding a glycine/tyrosine-rich keratin-associated protein of hair.";
RL J. Biol. Chem. 268:4511-4518(1993).
CC -!- FUNCTION: In the hair cortex, hair keratin intermediate filaments
CC are embedded in an interfilamentous matrix, consisting of hair
CC keratin-associated protein (KRTAP), which are essential for the
CC formation of a rigid and resistant hair shaft through their
CC extensive disulfide bond cross-linking with abundant cysteine
CC residues of hair keratins. The matrix proteins include the high-
CC sulfur and high-glycine-tyrosine keratins.
CC -!- SUBUNIT: Interacts with hair keratins.
CC -!- DEVELOPMENTAL STAGE: KAP6 proteins are first expressed in
CC differentiating hair shaft keratinocytes a considerable distance
CC above the proliferative zone of the follicle bulb.
CC -!- SIMILARITY: Belongs to the KAP6 Gly/Tyr-rich type II protein
CC family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
-----
DR EMBL; M95718; AAA31379.1; -
DR PIR; B45466; B45466.
KW Acetylation; Keratin; Multigene family.
FT INIT MET 0
FT MOD_RES 1 1 N-acetylcysteine.
SQ SEQUENCE 79 AA; 8143 MW; 3A45A75B029BFF4 CRC64;

Query Match    40.4%; Score 118.5; DB 1; Length 79;
Best Local Similarity 57.4%; Pred. No. 0.0003;
Matches 27; Conservative 1; Mismatches 10; Indels 9; Gaps 3;

QY 3 RGYGTSLLGSGVGPYGGSGFGF---SYGSGY-----GYGYGGYGVG 41
Db 11 RGYGCCGYGG-LGYGYGLGCGGLGSGYGGYRRLCGGCGGYGYG 56

```

```

RESULT 15
KL91_HUMAN
ID KL91_HUMAN STANDARD; PRT; 88 AA.
AC O81UB9;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Keratin associated protein 19-1 (High tyrosine-glycine keratin
DE associated protein 19.1) (Fragment).
GN Name=KRTAP19-1; Synonyms=KAP19.1, KRTAP19.1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Scalp;
RX MEDLINE=22359007; PubMed=12359730; DOI=10.1074/jbc.M206422200;
RA Rogers M.A., Langbein L., Winter H., Ehmann C., Praetzel S.,
RA Schweizer J.;
RT "Characterization of a first domain of human high glycine-tyrosine and
RT high sulfur keratin-associated protein (KAP) genes on chromosome
RT 21q22.1.";
RL J. Biol. Chem. 277:48993-49002(2002).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953; DOI=10.1038/35012518;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takegi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Saeaki T., Nagamine K., Mitsuuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Leirach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
CC -!- FUNCTION: In the hair cortex, hair keratin intermediate filaments
CC are embedded in an interfilamentous matrix, consisting of hair
CC keratin-associated protein (KRTAP), which are essential for the
CC formation of a rigid and resistant hair shaft through their
CC extensive disulfide bond cross-linking with abundant cysteine
CC residues of hair keratins. The matrix proteins include the high-
CC sulfur and high-glycine-tyrosine keratins.
CC -!- SUBUNIT: Interacts with hair keratins.
CC -!- TISSUE SPECIFICITY: Detected in the upper portion of the hair
CC cortex.
CC -!- SIMILARITY: Belongs to the KRTAP type 19 family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
-----
DR EMBL; AJ457067; CAD29723.1; -
DR EMBL; AP000567; -; NOT ANNOTATED_CDS.
DR Genbank; HGNC:18936; KRTAP19-1.
KW Keratin; Multigene family; Repeat.
FT NON_TER 1 1
FT DOMAIN 3 82 26 X 2 AA repeats of G-[YCGS].
SQ SEQUENCE 88 AA; 8789 MW; 9268103BFD016D38 CRC64;

Query Match    40.3%; Score 118; DB 1; Length 88;
Best Local Similarity 45.2%; Pred. No. 0.00037;
Matches 28; Conservative 3; Mismatches 9; Indels 22; Gaps 3;

```

QY 4 GYTSLLG-GSVGYPYG-----GSGFGSGGYGY-GYGYGYGY 41
Db 8 GLGYSCGFEGGLGYGYGCCGFCRRSGCGYGYGYGSGFGSGGYGYGSGFG 67
QY 42 GY 43
Db 68 GY 69

Search completed: April 1, 2005, 10:15:02
Job time : 25.2809 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2005, 09:31:32 ; Search time 237.953 seconds
(without alignments)
1054.491 Million cell updates/sec

Title: US-09-891-064A-2_COPY_33_522

Perfect score: 2590
Sequence: 1 MHVRPMLSQPAYSPYFPEDEI.....LKSKLSHIKMGVDYDROKT 490

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2590	100.0	522	1 OCLN_HUMAN	Q16625 homo sapien
2	2410.5	93.1	521	1 OCLN_CANFA	Q28269 canis famil
3	2410.5	93.1	560	2 O9N0W3	Q9N0W3 canis famil
4	2335.5	90.2	521	1 OCLN_MOUSE	Q61146 mus musculu
5	2326.5	89.8	523	2 Q6P6T5	Q6P6T5 rattus norv
6	2323.5	89.7	523	2 Q9Z303	Q9Z303 rattus norv
7	2152	83.1	449	2 Q6P6P8	Q6P6P8 bos tauris
8	1431	55.3	519	2 Q6G748	Q6G748 xenopus lae
9	1382	53.4	493	1 OCLN_XENLA	Q9PUN1 xenopus lae
10	1149	44.4	500	2 Q6NX99	Q6NX99 brachydanio
11	1144	44.2	492	2 Q7ZUE7	Q7ZUE7 brachydanio
12	1086.5	41.9	497	2 Q6NRR0	Q6NRR0 xenopus lae
13	1070.5	41.3	504	1 OCLN_CHICK	Q91049 gallus gall
14	1030.5	39.8	489	1 OCLN_POTTR	Q28793 potorous tr
15	625	24.1	121	2 Q97765	Q97765 sus scrofa
16	369	14.2	81	2 Q91XW5	Q91XW5 rattus norv
17	294	11.4	58	2 Q93M26	Q93M26 ovis aries
18	294	11.4	58	2 Q95M48	Q95M48 bos tauris
19	287.5	11.1	558	2 Q8N4S9	Q8N4S9 bos tauris
20	281.5	10.9	436	2 Q93LE8	Q93LE8 mus musculu
21	236.5	9.1	602	1 ELL_MOUSE	Q08856 mus musculu
22	229.5	8.9	621	1 ELL_HUMAN	P55199 homo sapien
23	224	8.6	309	2 Q72656	Q72656 homo sapien
24	220.5	8.5	633	2 Q6PEG4	Q6PEG4 brachydanio
25	219	8.5	640	1 ELL2_HUMAN	Q00472 homo sapien
26	178	6.9	190	2 Q922M9	Q922M9 mus musculu
27	169.5	6.5	140	2 Q80UJ4	Q80UJ4 mus musculu
28	165	6.4	395	1 ELL3_MOUSE	Q80V12 mus musculu
29	165	6.4	397	1 ELL3_HUMAN	Q9B65 homo sapien
30	155	6.0	457	2 Q96NM9	Q96NM9 homo sapien
31	147	5.7	575	2 Q8GN45	Q8GN45 escherichia

32	145	5.6	1059	2 Q9VW51	Q9VW51 drosophila
33	143	5.5	430	2 Q95RS9	Q95RS9 drosophila
34	142.5	5.5	627	1 K2C1_MOUSE	P04104 mus musculu
35	142	5.5	1060	2 Q95VE6	Q95VE6 drosophila
36	141	5.4	136	2 Q6DJ30	Q6DJ30 xenopus tro
37	138.5	5.3	576	1 CEAT7_ECOLI	Q47112 escherichia
38	136.5	5.3	671	2 Q6CUQ3	Q6CUQ3 kluyveromyc
39	136.5	5.3	671	2 Q70723	Q70723 kluyveromyc
40	133.5	5.2	219	2 Q9CX17	Q9CX17 mus musculu
41	133.5	5.2	562	2 Q9Z331	Q9Z331 mus musculu
42	133	5.1	529	2 Q919P5	Q919P5 rana cateb
43	133	5.1	563	1 K2CF_HUMAN	P48669 homo sapien
44	131	5.1	553	2 Q9Z332	Q9Z332 mus musculu
45	131	5.1	563	1 K2CA_HUMAN	P02538 homo sapien

ALIGNMENTS

RESULT 1
ID OCLN_HUMAN STANDARD; PRT; 522 AA.
AC Q16625; Q8N6K1;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Occludin.
GN Name=OCLN;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon carcinoma;
RX MEDLINE=96181088; PubMed=8601611; DOI=10.1083/jcb.133.1.43;
RA Ando-Akatsuka Y., Saitou M., Hirase T., Kishi M., Sakakibara A., Itoh M., Yonemura S., Furuse M., Tsukita S.;
RT "Interspecies diversity of the occludin sequence: cDNA cloning of human, mouse, dog, and rat-kangaroo homologues.";
RL J. Cell Biol. 133:43-47(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA van Itallie C.M., Fanning A.S., Anderson J.M.;
RL Submitted (MAY-1996) to the ENBL/GenBank/DBBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Fukasawa M., Toyota T., Yoshitsugu K., Yoshikawa T.;
RL "Genomic structure of occludin gene.";
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, and Lung;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan B., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Soares M.B., Bonaldo M.F., Casavant T.L., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Rubin G.M., Rubin G.M., Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzly D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield A.C., Kravinsky M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human

Db 510 KRMVGDYDRKK 520

RESULT 5

Q6P6T5 PRELIMINARY; PRT; 523 AA.

ID Q6P6T5

AC Q6P6T5

DT 05-JUL-2004 (TrEMBLrel. 27, Created)

DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)

DE Occludin.

DE Occludin.

GN Name=Occln;

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Prostate;

RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zesberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hong L.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalón D.K., Munz D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W., Butterfield Y.S.,
Krzywinski M.I., Skalska U., Smalick J., Schnerk A., Schein J.E.,
Jones S.-J., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]

RP SEQUENCE FROM N.A.

RC TISSUE=Prostate;

RA Strausberg R.;

RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.

RE EMBL; BC062037; AH62037.1; --

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0005923; C:tight junction; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR InterPro; IPR008253; MARVEL.

DR InterPro; IPR002958; Occludin.

DR InterPro; IPR010844; Occludin_ELL.

DR Pfam; PF01284; MARVEL; 1.

DR Pfam; PF07303; Occludin_ELL; 1.

DR PRINTS; PR01258; OCCLUDIN.

SQ SEQUENCE 523 AA; 59186 MW; A31E007B58AEAB55 CRC64;

Query Match 89.8%; Score 2326.5; DB 2; Length 523;
Best Local Similarity 88.6%; Pred. No. 2.2e-129;
Matches 436; Conservative 25; Mismatches 28; Indels 3; Gaps 3;

QY 1 MHVRPMLSPAFSPFPEDEILHFYKWTSPGPGVIRILSLMVICIAIPACVASTLAWDRG 60
DB 33 MHVRPMLSPAFSPFPEDEILHFYKWTSPGPGVIRILSLMVICIAIPACVASTLAWDR 92
QY 61 YGTSLGGSGVPYCGSGFGSGGY-GYGYGYGYYGTYTPDPAAGFLMAAFCFIA 119
DB 93 YGTGIFGGSMMYPY-GSGFGSYGGGFGGYGYGYGYGYTDPAAGFLMAAFCFIA 151
QY 120 ALVIFTVSIRSEMRTRYLSLVIIISAILGIWFIATIVYINGVNPTAQSSGLYSQ 179
DB 152 SLVIFTVSIRSEMRTRYLSLVIIISAILGIWFIATIVYINGVNPTAQSSGMYSQ 211

RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA	Rasak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA	Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA	Krzywinski M.I., Skalska U., Smalhus D.E., Schmerch A., Schein J.E.,
RA	Jones S.J., Marra M.A.
RT	"Generation and initial analysis of more than 15,000 full-length human
RT	and mouse cDNA sequences".
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN	[2].
RP	SEQUENCE FROM N.A.
RC	TISSUE=Whole body;
RA	Strausberg R.;
RL	Submitted (MAR-2003) to the EMBL/GenBank/DBAJ databases.
DR	EMBL; BC049304; AAH49304.1; --
DR	ZFIN; ZDR-GENE-040426-2685; zgc:56359.
DR	GO; GO:0016021; C:integral to membrane; IEA.
DR	GO; GO:0005923; C:tight junction; IEA.
DR	GO; GO:0005198; F:structural molecule activity; IEA.
DR	InterPro; IPR008253; Marvel.
DR	InterPro; IPR002958; Occludin.
DR	InterPro; IPR010844; Occludin_ELL.
DR	Pfam; PF01284; MARVEL; 1.
DR	Pfam; PF07303; Occludin_ELL; 1.
DR	PRINTS; PR01258; OCCLUDIN.
SQ	SEQUENCE 492 AA; 55173 MW; 4B92D6257361F2A5 CRC64;
Query Match	44.2%; Score 1144; DB 2; Length 492;
Best Local Similarity	46.4%; Pred. No. 1.4e-59;
Matches 228; Conservative	90; Mismatches 139; Indels 34; Gaps 11;
QY	9 QPAYSFPEDBIILHFYKWTSPGVIRILSMIIIVMCIAIFACVASTLAWDRGYGTSILGG 68
Db	: : : : : : : : : : : : : : : : : : : :
QY	24 QPAYSYPPDEFQHFFYRWTSPPGIKIMCVLSIIFCVGIFVCVASTLAWDTNAGAGFGT 83
Db	: : : : : : : : : : : : : : : : : : : :
QY	69 SVGYPYCGSGFGSGYG---GYG-----YGYG-YGYTDPPRAKGFLMAAPCFITAA 120
Db	: : : : : : : : : : : : : : : : : : : :
QY	84 NGGY-YGGSYAGSYSSGFGGTGYGMGAGSGFYGILGSQNDPROCKGFMIAMAITFTAL 142
Db	: : : : : : : : : : : : : : : : : : : :
QY	121 LVIPVTSVIRSEMSRTRYLVSIIVSAIGLMVFATIVVMGNVPTAQSSGSILYSQSI 180
Db	: : : : : : : : : : : : : : : : : : : :
QY	143 MVFIWISHQRVSQGRKFVLSIIIVSALLAFPFVIATIVLVTVYPWAQTSGSVQNQV 202
Db	: : : : : : : : : : : : : : : : : : : :
QY	181 YALCNQFYTPAATGLYYDVQLYHYCVVDPOBAIAIVLGFMIIIVAFALIIFPAVKTRKMD 240
Db	: : : : : : : : : : : : : : : : : : : :
QY	203 YSMCAAYQNPMQSGAFVNQYLXHYCVVDPOBAIALVLDFVVAALIIIMVFAIKTRQIN 262
Db	: : : : : : : : : : : : : : : : : : : :
QY	241 RYDKSNILLDKELHIYDEQ--PPNVEEVKNVSACTQDPVPSPPDYVERVDSPMAYSSNGK 298
Db	: : : : : : : : : : : : : : : : : : : :
QY	263 NYGKDNILMRVRKEFDQNSQPDVEDMWNVNGAPEGL-----LADYPVFKFSRNN 313
Db	: : : : : : : : : : : : : : : : : : : :
QY	299 VNDKRFPYESSYKSTPVPE-VVQELPLTSPVDDPQRPYSGGNFETPSKRAPAKRAGR 357
Db	: : : : : : : : : : : : : : : : : : : :
QY	314 LDDN----STSVDKPPLSESPEVILPVNSVP-----ISSGSEINSVGR-PKKRRAGR 362
Db	: : : : : : : : : : : : : : : : : : : :
QY	358 SKTEQDHETYDTTGDESCDELEDWIREFYPPITSQQOQLYKRNFTDTGLQEKYSIQSE 417
Db	: : : : : : : : : : : : : : : : : : : :
QY	363 PRTADGRDYDADYASSGDEL--DDDFSFEPPIVNTQERDDYKHLPDQHQEYKDLOAE 420
Db	: : : : : : : : : : : : : : : : : : : :
QY	418 LDEINKSLRLDKELDDRYESBEYMAAADENKLQVKSADYKSKNHCKOLKSLSH 477
Db	: : : : : : : : : : : : : : : : : : : :
QY	421 MDQINKLAERDLGDLQSGSPFLDAMEYNAIQOKRSGEYKQKKCKYLDKLNH 480
Db	: : : : : : : : : : : : : : : : : : : :
QY	478 IKKWVGDDYDRQ 488
Db	: : : : : : : : : : : : : : : : : : : :
QY	481 IKKMVSIDRR 491
Db	: : : : : : : : : : : : : : : : : : : :

Db 80 ---GGLG---GGGLSGYNGYGGYN-GYYGGLTNPRAANGFMAMAVLCFLVTLGL 133

Qy 124 FVTSVIRSEMGRTRYLSVIVSAAILGIMVFIATVYIMGVNPTA---OSSGSLYGSQI 180

Db 133 VIAGLSKASGARSRRFYLLVAVLSGLLAFVNLIASIVVGVNPRAGLGASSGSLYNNQ 192

Qy 181 YALCNQFTPAATGLYVDYLYHYCVDPQBAIAVLGFMIIIVAFALIIIPAVKTRKMD 240

Db 193 LMLCNQMSPVAGGI-MNQYLYHYCMVDPQBAIVLGVFLTVLLVCVYCPAQKTRHKI 251

Qy 241 RYDKSNILWKEHIYDEOPNNVEWVKVNS--AGTQDVPSPPSYDVRVDSPPMAYSSNGK 298

Db 252 KYGKNIWFDXP-LATAEGPNVEWVKVNSGDVGTQDETATLA-YSEKPIISPLT----- 303

Qy 299 VNDKRFYP--ESSY-KSTPPEVVOELPLTSPVDPDFRQPRYSSGGNFETPSKRAKRA 355

Db 304 ---SAFLPAQENGYGHSTP-----SSPSVPPPEGSPPEEKDGXSVSRPPAR---RGHR 351

Qy 356 GRSKKT--EODHYETDYTTGSCDELEL-BEDWIREYPPITSDQOROLYKRNFDFTGLQYK 412

Db 352 QRPRPTGLEESQYETDYTTAAESGEQNRDDWASLYPPIISDAIRQTYKAEFNNDLQRYK 411

Qy 413 SLOSELDEINKSLRLDKELDDYRESEYMAADEVNRLKOVKGSADYKSKKNHCKOLK 472

Db 412 ALCAEMDDIGTQLRQLSHELCLPEGLSURYOGVAEYNRLKDLKRSPEYQSKKLETOSLR 471

Qy 473 SKLSHIKRMVGDYDRQKT 490

Db 472 DLKCHIKRMVGDYDQSR 489

RESULT 15

O97766 PRELIMINARY; PRT; 121 AA.

ID O97766 AC O97766; AC O97766; PRELIMINARY; PRT; 121 AA.

DT 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DE 01-WAR-2004 (TREMBLrel. 26, Last annotation update)

DE Occludin (Fragment).

OS Sus acrofa (Pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

OX NCBI_TaxID=9823;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=97388360; PubMed=9247194;

RA Hirase T., Staddon J.M., Saitou M., Ando-Akatsuka Y., Itoh M., Furuse M., Fujimoto K., Tsukita S., Ruben L.L.;

RT "Occludin as a possible determinant of tight junction permeability in endothelial cells."

RL J. Cell Sci. 110:1603-1613 (1997).

DR EMBL; U79554; AAD00349.1; -

DR InterPro; IPR010844; Occludin_ELL.

DR Pfam; PF07303; Occludin_ELL; I.

FT NON TER 1

FT NON TER 121

SQ SEQUENCE 121 AA; 14433 MW; 6B37FD57A5F647A6 CRC64;

Query Match 24.1%; Score 625; DB 2; Length 121;

Best Local Similarity 96.7%; Pred. No. 1.3e-29;

Matches 117; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 366 YETDVTGSCDELEEDWIREYPPITSDQOROLYKRNFDFTGLQYKSLQELDEINKEL 425

Db 1 YETDVTGSCDELEEDWIREYPPITSDQOROLYKRNFDFTGLQYKSLQELDEINKEL 60

Qy 426 SRLDKELDDYRESEYMAADEVNRLKOVKGSADYKSKKNHCKOLKSLSHIKWGVGY 485

Db 61 SRLDKELDDYRESEYMAADEVNRLKOVKGSADYKSKKNHCKOLKSLSHIKWGVGY 120

Qy 486 D 486

Db 121 D 121

Search completed: April 1, 2005, 10:15:01
Job time : 240.953 secs

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	293	100.0	522	9	US-09-891-064A-2	Sequence 2, Appli
2	293	100.0	522	14	US-10-176-847-104	Sequence 104, App
3	293	100.0	522	15	US-10-601-953-900	Sequence 900, App
4	141	48.1	24	9	US-09-891-064A-4	Sequence 4, Appli
5	128	43.7	24	9	US-09-891-064A-3	Sequence 3, Appli
6	117	39.9	135	16	US-10-437-963-129038	Sequence 129038,
7	116.5	39.8	174	16	US-10-437-963-174945	Sequence 174945,
8	115.5	39.4	140	16	US-10-767-701-36790	Sequence 36790, A
9	114.5	39.1	336	15	US-10-425-114-67770	Sequence 67770, A
10	110.5	37.7	416	15	US-10-425-114-64457	Sequence 64457, A
11	108.5	37.0	134	14	US-10-280-114-13	Sequence 13, Appli
12	108	36.9	174	15	US-10-425-114-62338	Sequence 62338, A
13	106.5	36.3	154	14	US-10-226-489-16	Sequence 16, Appli

```
US-10-176-847-104
; Sequence 104, Application US/10176847
; Publication No. US20030068636A1
; GENERAL INFORMATION:
; APPLICANT: Veiby, Petter Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
; TITLE OF INVENTION: AND OVARIAN CANCER
; FILE REFERENCE: MRI-039
; CURRENT APPLICATION NUMBER: US/10/176,847
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 104
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-176-847-104

Query Match      100.0%; Score 293; DB 14; Length 522;
Best Local Similarity 100.0%; Pred. No. 6.1e-24;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WDRGYGTSLLGGSGVGYPGYGGSGFGSGYGYGYGYGYGYGYGYGYTDPRAAK 50
Db 89 WDRGYGTSLLGGSGVGYPGYGGSGFGSGYGYGYGYGYGYGYGYGYTDPRAAK 138

RESULT 3
US-10-601-953-900
; Sequence 900, Application US/10601953
; Publication No. US20040077540A1
; GENERAL INFORMATION:
; APPLICANT: Quay, Steven C.
; TITLE OF INVENTION: Compositions And Methods For Modulating Physiology Of Epithelial
; TITLE OF INVENTION: Junctional Adhesion Molecules For Enhanced Mucosal Delivery Of
; TITLE OF INVENTION: Therapeutic Compounds
; FILE REFERENCE: 02-0305
; CURRENT APPLICATION NUMBER: US/10/601,953
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 60/392,512
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 900
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 900
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
; US-10-601-953-900

Query Match      100.0%; Score 293; DB 15; Length 522;
Best Local Similarity 100.0%; Pred. No. 6.1e-24;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WDRGYGTSLLGGSGVGYPGYGGSGFGSGYGYGYGYGYGYGYGYGYTDPRAAK 50
Db 89 WDRGYGTSLLGGSGVGYPGYGGSGFGSGYGYGYGYGYGYGYGYGYTDPRAAK 138

RESULT 4
US-09-891-064A-4
; Sequence 4, Application US/09891064A
; Patent No. US20020082391A1
; GENERAL INFORMATION:
; APPLICANT: James M. Anderson
; APPLICANT: Christina M. Van Itallie
; TITLE OF INVENTION: Human Occludin, Its Uses and Enhancement of Drug
; TITLE OF INVENTION: Absorption Using Occludin Inhibitors
; FILE REFERENCE: OCR-754.CIP
; CURRENT APPLICATION NUMBER: US/09/891,064A
; CURRENT FILING DATE: 2001-06-25
```

```
; PRIOR APPLICATION NUMBER: US 09/142,732
; PRIOR FILING DATE: 1998-09-15
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: MS DOS
; SEQ ID NO 4
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: peptide
; OTHER INFORMATION: construct used in experiments
US-09-891-064A-4

Query Match      48.1%; Score 141; DB 9; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.3e-08;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 SYGSGYGYGYGYGYGYGYGYTDP 47
Db 2 SYGSGYGYGYGYGYGYGYTDP 24

RESULT 5
US-09-891-064A-3
; Sequence 3, Application US/09891064A
; Patent No. US20020082391A1
; GENERAL INFORMATION:
; APPLICANT: James M. Anderson
; APPLICANT: Christina M. Van Itallie
; TITLE OF INVENTION: Human Occludin, Its Uses and Enhancement of Drug
; TITLE OF INVENTION: Absorption Using Occludin Inhibitors
; FILE REFERENCE: OCR-754.CIP
; CURRENT APPLICATION NUMBER: US/09/891,064A
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 09/142,732
; PRIOR FILING DATE: 1998-09-15
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: MS DOS
; SEQ ID NO 3
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: peptide
; OTHER INFORMATION: construct used in experiments
US-09-891-064A-3

Query Match      43.7%; Score 128; DB 9; Length 24;
Best Local Similarity 100.0%; Pred. No. 3.5e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 DRGYGTSLLGGSGVGYPGYGGSGFG 24
Db 2 DRGYGTSLLGGSGVGYPGYGGSGFG 24

RESULT 6
US-10-437-963-129038
; Sequence 129038, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
```

```
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 129038
; LENGTH: 135
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_31333C.1.pap
US-10-437-963-129038

Query Match          39.9%; Score 117; DB 16; Length 135;
Best Local Similarity 46.4%; Pred. No. 2.8e-05;
Matches 26; Conservative 1; Mismatches 11; Indels 18; Gaps 2;

QY 4 GYGTSLGSGVGP-----YGGSGFGSYGSGYGYGYGYGG 42
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 73 GYGRPGYGGYHPGYGGYGGYGRGYGGYGGSG-GYGGYGGYGGGGYGG 127

RESULT 7
US-10-437-963-174945
; Sequence 174945, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 174945
; LENGTH: 174
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_72833C.1.pap
US-10-437-963-174945

Query Match          39.8%; Score 116.5; DB 16; Length 174;
Best Local Similarity 59.5%; Pred. No. 4.1e-05;
Matches 25; Conservative 2; Mismatches 12; Indels 3; Gaps 2;

QY 4 GYGTSLGSGVGP--YGGSGF-GSYGSGYGYGYGYGYGG 42
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108 GYGPYGGYGGSPGYGGSGYGGYGGYGGYGGYGGYGG 149

RESULT 8
US-10-767-701-36790
; Sequence 36790, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 36790
; LENGTH: 140
; TYPE: PRT
; ORGANISM: Sorghum bicolor
```

```
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(140)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-Cl19405_1.pap
US-10-767-701-36790

Query Match          39.4%; Score 115.5; DB 16; Length 140;
Best Local Similarity 42.6%; Pred. No. 4.3e-05;
Matches 26; Conservative 1; Mismatches 13; Indels 21; Gaps 2;

QY 4 GYGTSLGSGVGPYGGSGFGS-----YGGSGYGYGYGYGYG---G 42
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70 GYGPSYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGG 129

QY 43 Y 43
Db 130 Y 130

RESULT 9
US-10-425-114-67770
; Sequence 67770, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 67770
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LTB3632-012-C2_FLI.pap
US-10-425-114-67770

Query Match          39.1%; Score 114.5; DB 15; Length 336;
Best Local Similarity 52.2%; Pred. No. 0.00013;
Matches 24; Conservative 4; Mismatches 11; Indels 7; Gaps 2;

QY 4 GYGTSLGSGVGPYGG-----SGFG-SYGGSGYGYGYGYGYGG 42
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 203 GAGSGSGGGGGYGGYGGYRGAAGYDGGAGAGYGGYGGYGG 248

RESULT 10
US-10-425-114-64457
; Sequence 64457, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 64457
```

```
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3960-010-H9_FLI.pep
US-10-425-114-64457

Query Match          37.7%; Score 110.5; DB 15; Length 416;
Best Local Similarity 43.9%; Pred. No. 0.00042;
Matches 29; Conservative 4; Mismatches 14; Indels 19; Gaps 5;
Qy 3 RGYGTSL-LGSSVGYP-----YGGSGFGSGYSGYGYG---GYGYGYG--YGG--Y 43
Db 299 RGSQVSVNSGGYGYGGGYSRAAAAYGSTAYGAYGCGYGYGATSGYSGYGSAYGGSMY 358
Qy 44 TDPRAA 49
Db 359 GGPYGA 364

RESULT 11
US-10-280-114-13
; Sequence 13, Application US/10280114
; Publication No. US20030170257A1
; GENERAL INFORMATION:
; APPLICANT: Trimmell, Adama Roseanne
; APPLICANT: Paesen, Guido Christiaan
; APPLICANT: Nuttall, Patricia Anne
; TITLE OF INVENTION: Vaccine Comprising A Tick Cement Protein
; FILE REFERENCE: 2488-1-007
; CURRENT APPLICATION NUMBER: US/10/280,114
; PRIOR FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: PCT/GB01/01834
; PRIOR FILING DATE: 2001-04-25
; PRIOR FILING DATE: 2000-04-25
; PRIOR FILING DATE: 2000-04-25
; PRIOR FILING DATE: 2000-11-23
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: SeqWin99, version 1.02
; SEQ ID NO 13
; LENGTH: 134
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Clone 64
US-10-280-114-13

Query Match          37.0%; Score 108.5; DB 14; Length 134;
Best Local Similarity 52.2%; Pred. No. 0.00024;
Matches 24; Conservative 4; Mismatches 13; Indels 5; Gaps 3;
Qy 3 RGYG--TSLLGSGV--YYPGG-SGFGSGYSGYGYGYGYGYGY 43
Db 48 RGYPLGTASIGGEVGARLPYGGYGGYGGYGGYDQGFSGAYGGY 93

RESULT 12
US-10-425-114-62338
; Sequence 62338, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
```

```
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 62338
; LENGTH: 174
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4730-008-E7_FLI.pep
US-10-425-114-62338

Query Match          36.9%; Score 108; DB 15; Length 174;
Best Local Similarity 53.2%; Pred. No. 0.00034;
Matches 25; Conservative 0; Mismatches 12; Indels 10; Gaps 2;
Qy 2 DRGYGTSLLGSGVY-----PYGGSGFGSGYSGYGYGYGYGYGG 42
Db 129 DRGYG-----GDRGYGGGGDRGYGGGGDRGYGGGGGGGGG 171

RESULT 13
US-10-226-489-16
; Sequence 16, Application US/10226489
; Publication No. US20030078390A1
; GENERAL INFORMATION:
; APPLICANT: Guido Christiaan Paesen, Patricia Anne Nuttall
; TITLE OF INVENTION: Tissue Cement
; FILE REFERENCE: 2488-1-001
; CURRENT APPLICATION NUMBER: US/10/226,489
; CURRENT FILING DATE: 2002-08-23
; PRIOR FILING DATE: 2002-08-23
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 154
; TYPE: PRT
; ORGANISM: R. appendiculatus
US-10-226-489-16

Query Match          36.3%; Score 106.5; DB 14; Length 154;
Best Local Similarity 37.9%; Pred. No. 0.00045;
Matches 25; Conservative 4; Mismatches 12; Indels 25; Gaps 3;
Qy 3 RGYG--TSLLGSGV-----GYPYGG-SGFGSGYSGYGYGYGYG 37
Db 48 RGYPLGTASIGGEVGARLGGRAGVGVSYSYGYGYSWGYPYGGYGGYGGYDQGF 107
Qy 38 YGYGGY 43
Db 108 SAYGGY 113

RESULT 14
US-10-280-114-17
; Sequence 17, Application US/10280114
; Publication No. US20030170257A1
; GENERAL INFORMATION:
; APPLICANT: Trimmell, Adama Roseanne
; APPLICANT: Paesen, Guido Christiaan
; APPLICANT: Nuttall, Patricia Anne
; TITLE OF INVENTION: Vaccine Comprising A Tick Cement Protein
; FILE REFERENCE: 2488-1-007
; CURRENT APPLICATION NUMBER: US/10/280,114
; CURRENT FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: PCT/GB01/01834
; PRIOR FILING DATE: 2001-04-25
; PRIOR FILING DATE: 2000-04-25
; PRIOR FILING DATE: 2000-04-25
; PRIOR FILING DATE: 2000-11-23
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: SeqWin99, version 1.02
; SEQ ID NO 17
; LENGTH: 154
```

```

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: 64P protein fragment
US-10-280-114-17

Query Match      36.3%; Score 106.5; DB 14; Length 154;
Best Local Similarity 37.9%; Pred. No. 0.00045;
Matches 25; Conservative 4; Mismatches 12; Indels 25; Gaps 3;

QY 3 RGYG--TSLGGSV-----GYPYGG-SGFGSYGSGYGYGYG 37
Db 48 RGYPLGTASIGGEVGARLGAGVGVSSYGYGYPGSGYGYGYGYG 107
QY 38 YGYGGY 43
Db 108 SAYGGY 113

RESULT 15
US-10-425-114-64513
; Sequence 64513, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 64513
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3066-032-B8_FLI.pep
US-10-425-114-64513

Query Match      36.3%; Score 106.5; DB 15; Length 336;
Best Local Similarity 41.4%; Pred. No. 0.00094;
Matches 29; Conservative 4; Mismatches 14; Indels 23; Gaps 5;

QY 3 RGYGTSL-----GGSVGYP-----YGGSGFGSYGSGYGYG 41
Db 215 RSGSGSGYSGNSGGYGGYRSTAAAHYGSTAYAGRGYGYGTTAGYGLGYSAYG 274
QY 42 G--YTDPRAA 49
Db 275 GMYGGPYGA 284

Search completed: April 1, 2005, 10:03:20
Job time : 135.856 secs

```

BEST AVAILABLE COPY

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2005, 09:36:47 ; Search time 5.32995 Seconds
(without alignments)
902.605 Million cell updates/sec

Title: US-09-891-064a-2_COPY_89_138

Perfect score: 293
Sequence: 1 WDRGYGTSLLGSGVGYGG.....GYGYGYGYGYGYTDPRAAK 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	293	100.0	522	2 G02533	occludin - human
2	118.5	40.4	80	2 B45466	glycine/tyrosine-r
3	116	38.6	180	2 JC7876	prion protein homo
4	112	38.2	210	2 C87305	hypothetical prote
5	112	38.2	404	2 B4745	probable RNA-bind
6	110.5	37.7	102	2 T20557	hypothetical prote
7	109	37.2	504	2 A49467	occludin - chicken
8	108.5	37.0	155	1 KRCHS	keratin, scale - c
9	108	36.9	168	2 C24255	chorion class B pr
10	107.5	36.7	119	2 S04513	chorion class B pr
11	106.5	36.3	227	2 T15772	hypothetical prote
12	106	36.2	129	2 A24255	chorion class A pr
13	105	35.8	91	2 S01421	chorion class B pr
14	105	35.8	151	2 S04515	chorion class B pr
15	105	35.8	353	2 B34504	heterogeneous nucl
16	104	35.5	75	2 A89016	protein B0213.6 [i
17	104	35.5	183	1 KNRZG2	glycine-rich cell
18	104	35.5	200	2 S10334	glycine-rich prote
19	103	35.2	186	2 S28021	rab18 protein - Ar
20	102.5	35.0	53	2 S12520	core protein A1 -
21	102	34.8	386	2 S51436	probable membrane
22	101.5	34.6	126	2 S04514	chorion class B pr
23	101.5	34.6	345	1 B41732	heterogeneous nucl
24	101	34.5	177	2 A31204	eggshell protein p
25	100.5	34.3	69	2 B93016	protein B0213.5 [i
26	100.5	34.3	165	2 T03583	glycine-rich RNA-b
27	100	34.1	301	2 JW0079	heterogeneous nucl
28	100	34.1	371	2 E88633	protein F56B3.1 [i
29	100	34.1	526	1 KRBOVI	keratin, 54K type

ALIGNMENTS

RESULT 1

G02533

occludin - human

C:Species: Homo sapiens (man)

C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 09-Jul-2004

C:Accession: G02533

R:Van Itallie, C.M.

submitted to the EMBL Data Library, April 1996

A:Reference number: H01403

A:Accession: G02533

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-522 <VAN>

A:Cross-references: UNIPROT:Q16625; EMBL:U53823; NID:gl322281; PIDN:AAB00195.1; PID:gl332

C:Superfamily: occludin

Query Match 100.0%; Score 293; DB 2; Length 522;

Best Local Similarity 100.0%; Pred. No. 6.6e-22;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WDRGYGTSLLGSGVGYGGSGFGSGVGYGYGYGYGYGYTDPRAAK 50

Db 89 WDRGYGTSLLGSGVGYGGSGFGSGVGYGYGYGYGYGYTDPRAAK 138

RESULT 2

B45466

glycine/tyrosine-rich keratin-associated hair protein KAP6.1 - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)

C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004

C:Accession: B45466

R:Frattini, A.; Powell, B.C.; Rogers, G.E.

J. Biol. Chem. 268, 4511-4518, 1993

A:Title: Sequence, expression, and evolutionary conservation of a gene encoding a glycin

A:Reference number: A45466; MUID:93179467; PMID:7680040

A:Accession: B45466

A:Molecule type: DNA

A:Residues: 1-80 <FRA>

A:Cross-references: UNIPROT:Q02957; GB:M95718; NID:gl65446; PIDN:AAA31379.1; PID:gl65447

A>Note: sequence extracted from NCBI backbone (NCBIN:125967; NCBIIP:125968)

Query Match 40.4%; Score 118.5; DB 2; Length 80;

Best Local Similarity 57.4%; Pred. No. 2.3e-05;

Matches 27; Conservative 1; Mismatches 10; Indels 9; Gaps 3;

Qy 3 RGYGTSLLGSGVGYGGSGFG----SYGSGY-----GYGYGYGYGYG 41

Db 12 RGYGCGGYG-LGYGYGSLGGLGSLYGGCYRRLLGCGYGGYGYGYG 57

RESULT 3

JC7876

prion protein homolog - Fugu rubripes
C:Species: Fugu rubripes
C>Date: 09-Dec-2002 #sequence_revision 09-Dec-2002 #text_change 09-Jul-2004
C:Accession: JG7876
R:Suzuki, T.; Kurokawa, T.; Hashimoto, H.; Sugiyama, M.
Biochem. Biophys. Res. Commun. 294, 912-917, 2002
A:Title: cDNA sequence and tissue expression of Fugu rubripes prion protein-like: A candidate
A:Reference number: JG7876; MUID:22057338; PMID:12061771
A:Accession: JG7876
A:Molecule type: mRNA
A:Residues: 1-180 <SUZ>
A:CROSS-references: UNIPROT:O8J1J1; DDBJ:AB084918
C:Comment: This protein is a real orthologue of human prion protein and plays some roles
C:Genetics:
A:Gene: prp

Query Match 39.6%; Score 116; DB 2; Length 180;
Best Local Similarity 59.1%; Pred. No. 8.8e-05;
Matches 26; Conservative 2; Mismatches 12; Indels 4; Gaps 3;

Qy 4 GYGTSLGGSGVGYPGSGFGSGYSGYGYGYGYG--GGYTD 45
Db 97 GYGYGGYGG-YGMGYGGY-GGYGGYGGYGRGYGPRGGYSN 138

RESULT 4
C87305
hypothetical protein CC0452 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C:Accession: C87305
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBooy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: C87305
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-210 <STO>
A:CROSS-references: UNIPROT:Q9AAZ0; GB:AE005673; NID:g13421623; PIDN:AAK22439.1; GSPDB:G
C:Genetics:
A:Gene: CC0452

Query Match 38.2%; Score 112; DB 2; Length 210;
Best Local Similarity 65.5%; Pred. No. 0.00025;
Matches 19; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 12 GSVGYPYGGSGFGSGYSGYGYGYGYGY 40
Db 115 GKGGWYGGPGYDAYGPGYGYDYGYGY 143

RESULT 5
B84745
probable RNA-binding protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: B84745
R:Lin, X.; Kaul, S.; Rounalev, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vankken, S.E.; Unayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: B84745
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-404 <STO>
A:CROSS-references: UNIPROT:O22791; GB:AE002093; NID:g2459445; PIDN:AA80680.1; GSPDB:GN
C:Genetics:
A:Gene: At2g33410

A:Map position: 2
C:Superfamily: helix-destabilizing protein; ribonucleoprotein repeat homology

Query Match 38.2%; Score 112; DB 2; Length 404;
Best Local Similarity 44.9%; Pred. No. 0.00047;
Matches 31; Conservative 3; Mismatches 5; Indels 30; Gaps 6;

Qy 4 GYGTSLGGSGVGY-----PYGSGFGSGYSGYGY---GYGYGYGY 40
Db 210 GYGG--GGS-GYGRVDSNRYPQNTGSGYPPYGGG---YGTGYGSGNGVGYG-GF 262

Qy 41 GGYTDPRAA 49
Db 263 GGYGNPAGA 271

RESULT 6
T20557
hypothetical protein F07H5.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T20557
R:Steward, C.
submitted to the EMBL Data Library, December 1995
A:Reference number: Z19292
A:Accession: T20557
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-102 <WIL>
A:CROSS-references: UNIPROT:Q19176; EMBL:Z68314; PIDN:CAA92658.1; GSPDB:GN000020; CESP:F0
A:Experimental source: clone F07H5
C:Genetics:
A:Gene: CESP:F07H5.6
A:Map position: 2
A:Introns: 33/1; 85/1
C:Superfamily: Arabidopsis glycine-rich protein 3

Query Match 37.7%; Score 110.5; DB 2; Length 102;
Best Local Similarity 50.0%; Pred. No. 0.00018;
Matches 26; Conservative 1; Mismatches 12; Indels 13; Gaps 4;

Qy 4 GYGTSL-----LGGSGVGYPGSGFGSGYSGY-GYGY-GYGYGYGY 43
Db 27 GYGLGLGLCGGFGSGGGYGLGYGG-GFGYGGFGGYGLGYGYGY 77

RESULT 7
A49467
occludin - chicken
C:Species: Gallus gallus (chicken)
C>Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004
C:Accession: A49467
R:Furuse, M.; Hirase, T.; Itoh, M.; Nagafuchi, A.; Yonemura, S.; Tsukita, S.; Tsukita, S.;
J. Cell Biol. 123, 1777-1788, 1993
A:Title: Occludin: a novel integral membrane protein localizing at tight junctions.
A:Reference number: A49467; MUID:94103332; PMID:8276896
A:Accession: A49467
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-504 <FUR>
A:CROSS-references: UNIPROT:Q91049; GB:D21837; NID:g464148; PIDN:BAA04865.1; PID:g540494
C:Superfamily: occludin
C:Keywords: liver; membrane protein

Query Match 37.2%; Score 109; DB 2; Length 504;
Best Local Similarity 55.8%; Pred. No. 0.0011;
Matches 24; Conservative 1; Mismatches 14; Indels 4; Gaps 2;

Qy 1 WDRCYGTSLGGSGVGYGG-SGFGSGYSGYGYGYGYGG 42
Db 80 WDYG---LGGAYGTGLGGFYGSNYGSGLSYSGYGYGG 119

RESULT 8

KRCHS
keratin, scale - chicken
N:Alternate names: beta keratin
C:Species: Gallus gallus (chicken)
C>Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 09-Jul-2004
C:Accession: A02851; B60136; I50168
R:Gregg, K.; Wilton, S.D.; Parry, D.A.D.; Rogers, G.E.
EMBO J. 3, 175-178, 1984
A:Title: A comparison of genomic coding sequences for feather and scale keratins: structure
A:Reference number: A09088; MUID:84158528; PMID:6200321
A:Accession: A02851
A:Molecule type: DNA
A:Residues: 1-155 <GRE>
A:Cross-references: UNIPROT:P04459; GB:X00315; NID:G63548; PIDN:CAA25084.1; PID:G63549
R:Wilton, S.D.; Crocker, L.A.; Rogers, G.E.
Biochim. Biophys. Acta 824, 201-208, 1985
A:Title: Isolation and characterisation of keratin mRNA from the scale epidermis of the
A:Reference number: A60136; MUID:85122780; PMID:2578918
A:Accession: B60136
A:Molecule type: mRNA
A:Residues: 90-155 <WIL>
A:Cross-references: GB:M25642; NID:G212231; PIDN:AAA48931.1; PID:G212232
A:Note: This mRNA was designated clone CSK12. See entry A60136 for clone CSK9
R:Sawyer, R.H.; Shames, R.B.
Curr. Top. Dev. Biol. 22, 235-253, 1987
A:Title: Expression of beta-keratin genes during development of avian skin appendages.
A:Reference number: I50168; MUID:88003561; PMID:2443313
A:Accession: I50168
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 137-155 <SAW>
A:Cross-references: GB:M28422; NID:G211271; PIDN:AAA48631.1; PID:G211272
C:Superfamily: feather keratin
C:Keywords: duplication; epidermis; fibrous protein; horn; integument; tandem repeat
F:70-128/Region: 13-residue repeats (G-G-S-S-L-G-Y-G-G-L-Y-G-Y)

Query Match 37.0%; Score 108.5; DB 1; Length 155;
Best Local Similarity 50.9%; Pred. No. 0.00043;
Matches 27; Conservative 0; Mismatches 15; Indels 11; Gaps 3;

QY 4 GYGTSLGSGVYPGSGFGSGYGGYGVG-----YGYGYGVGYGTP 46
DB 88 GYGGSSLYGGLYGVGGSLG-YGGLYGVGGSLYGVGLYGVGRSYGSGYCS 139

RESULT 9
C24255
chorion class B protein L1 precursor - silkworm
C:Species: Bombyx mori (silkworm)
C>Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 09-Jul-2004
C:Accession: C24255; D24255; S05914
R:Sporel, N.; Nguyen, H.T.; Kafatos, F.C.
J. Mol. Biol. 190, 23-35, 1986
A:Title: Gene regulation and evolution in the chorion locus of Bombyx mori. Structural a
A:Reference number: A92929; MUID:87060979; PMID:3023635
A:Accession: C24255
A:Molecule type: DNA
A:Residues: 1-168 <SPO>
A:Cross-references: UNIPROT:P08827; GB:X15557; GB:X04028; GB:X04029; GB:X04030; GB:X0403
A:Accession: D24255
A:Molecule type: DNA
A:Residues: 1-18, 'V', 20, 'S', 22-27, 35, 'R', 37-50, 'L', 52-114, 'D', 116-130, 'F', 132, 'Y', 134-13
A:Cross-references: GB:X15557; GB:X04028; GB:X04029; GB:X04030; GB:X04031; GB:X04032; GB
R:Sporel, N.A.; Nguyen, H.T.; Eickbush, T.H.; Kafatos, F.C.
J. Mol. Biol. 209, 1-19, 1989
A:Title: Gene evolution and regulation in the chorion complex of Bombyx mori. Hybridizat
A:Reference number: S05913; MUID:90040707; PMID:2810362
A:Accession: S05914
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-17 <SP3>
A:Cross-references: EMBL:X15558; NID:G5771; PID:G5774

A:Note: this sequence was submitted to the EMBL Data Library, Jun-1989
C:Genetics:
A:Map position: 2
C:Superfamily: chorion class A protein pc292
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-168/Product: chorion class B protein L1 #status predicted <MAT>

Query Match 36.9%; Score 108; DB 2; Length 168;
Best Local Similarity 54.0%; Pred. No. 0.00051;
Matches 27; Conservative 2; Mismatches 11; Indels 10; Gaps 4;

QY 4 GYGTSLLG---GSVGYVP-YGSGPQS-YGSGYGVY-----GYGYGYGY 43
DB 110 GTGEILYCGNGAVGITREGGLGFGAGYGGYGLGYYGGYGLGYYGY 159

RESULT 10

S04513
chorion class B protein (clone M2807) - silkworm (fragment)
C:Species: Bombyx mori (silkworm)
C>Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 09-Jul-2004
C:Accession: S04513
R:Tsitilou, S.G.; Rodakis, G.C.; Alexopoulos, M.; Kafatos, F.C.; Ito, K.; Iatrou, K.
EMBO J. 2, 1845-1852, 1983
A:Title: Structural features of B family chorion sequences in the silkworm Bombyx mori.
A:Reference number: S01420; MUID:84057707; PMID:6571700
A:Accession: S04513
A:Molecule type: mRNA
A:Residues: 1-119 <TSI>
A:Cross-references: UNIPROT:P08914; EMBL:X12838; NID:G5813; PIDN:CAA1323.1; PID:G665607
C:Superfamily: chorion class A protein pc292

Query Match 36.7%; Score 107.5; DB 2; Length 119;
Best Local Similarity 52.1%; Pred. No. 0.00042;
Matches 25; Conservative 4; Mismatches 10; Indels 9; Gaps 4;

QY 2 DRGYGTSLLG---GSVGYPYGGSGFG-SYGSY-GYGYGYGYGYGY 43
DB 67 DYCGNGAVGITREGGLGY---GAGYGDYGLYGGYGGYGLGYYGY 111

RESULT 11

T15772
hypothetical protein C34F11.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T15772
R:Bentley, D.
submitted to the EMBL Data Library, January 1996
A:Description: The sequence of C. elegans cosmid C34F11.
A:Reference number: Z18400
A:Accession: T15772
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-227 <BEN>
A:Cross-references: UNIPROT:Q18464; EMBL:U46753; NID:G1166623; PID:G1166629; PIDN:AAA857
C:Genetics:
A:Gene: CESP:C34F11.8
A:Introns: 33/1; 73/3; 110/2; 189/1

Query Match 36.3%; Score 106.5; DB 2; Length 227;
Best Local Similarity 46.9%; Pred. No. 0.00096;
Matches 23; Conservative 1; Mismatches 14; Indels 11; Gaps 2;

QY 5 YGTSLGSGVY--PYGG-----SGFGSYGSGYGYGYGYGYGG 42
DB 144 YGYSRYGSGYGVYGVYSGYSGYSGYSGYSGYSGYSGYSGYSGY 192

RESULT 12

A24255
chorion class A protein L1 precursor - silkworm
C:Species: Bombyx mori (silkworm)

C>Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 09-Jul-2004
 C:Accession: A24255
 R:Spoerel, N.; Nguyen, H.T.; Kafatos, F.C.
 J. Mol. Biol. 190, 23-35, 1986
 A:Title: Gene regulation and evolution in the chorion locus of Bombyx mori. Structural a
 A:Reference number: A32929; MUID:87060979; PMID:3023635
 A:Accession: A24255
 A:Molecule type: DNA
 A:Residues: 1-129 <SPO>
 A:Cross-references: UNIPROT:P08826; GB:X15557; GB:X04028; GB:X04029; GB:X04030; GB:X0403
 C:Superfamily: chorion class A protein pc292
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-129/Product: chorion class A protein L11 #status predicted <MAT>

Query Match 36.2%; Score 106; DB 2; Length 129;
 Best Local Similarity 60.5%; Pred. No. 0.00064;
 Matches 26; Conservative 1; Mismatches 12; Indels 4; Gaps 4;

OY 3 RGYGTSLLGSGVGPYG-GSGFGSGVGYG-YGYGYGGY 43
 26 RGLGCGCGGGLG-CYGLGYGLGGYGGGYGGY-YGGY 66

Db 26 RGLGCGCGGGLG-CYGLGYGLGGYGGGYGGY-YGGY 66

RESULT 13
 S01421
 chorion class B protein (clone M3A5) - silkworm (fragment)
 C:Species: Bombyx mori (silkworm)
 C>Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 09-Jul-2004
 A:Accession: S01421
 R:Tsitilou, S.G.; Rodakis, G.C.; Alexopoulos, M.; Kafatos, F.C.; Ito, K.; Iatrou, K.
 EMBO J. 2, 1845-1852, 1983
 A:Title: Structural features of B family chorion sequences in the silkmooth Bombyx mori,
 A:Reference number: S01420; MUID:84057707; PMID:6571700
 A:Accession: S01421
 A:Molecule type: mRNA
 A:Residues: 1-91 <TSI>
 A:Cross-references: UNIPROT:P08915; EMBL:X12840; NID:95816; PIDN:CAA31325.1; PID:95817
 C:Superfamily: chorion class A protein pc292

Query Match 35.8%; Score 105; DB 2; Length 91;
 Best Local Similarity 52.0%; Pred. No. 0.00058;
 Matches 26; Conservative 3; Mismatches 11; Indels 10; Gaps 4;

OY 4 GYGTSLLG---GSGVGP-YGSGFGS-YGSGVGYG-----GYGYGGY 43
 33 GIGBILYCGCGAVGITREGGLYGAGYGGYGLGYGGYGGY 82

Db 33 GIGBILYCGCGAVGITREGGLYGAGYGGYGLGYGGYGGY 82

RESULT 14
 S04515
 chorion class B protein precursor (clone 410) - silkworm (fragment)
 C:Species: Bombyx mori (silkworm)
 C>Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 09-Jul-2004
 A:Accession: S04515
 R:Tsitilou, S.G.; Rodakis, G.C.; Alexopoulos, M.; Kafatos, F.C.; Ito, K.; Iatrou, K.
 EMBO J. 2, 1845-1852, 1983
 A:Title: Structural features of B family chorion sequences in the silkmooth Bombyx mori,
 A:Reference number: S01420; MUID:84057707; PMID:6571700
 A:Accession: S04515
 A>Status: not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-151 <TSI>
 A:Cross-references: UNIPROT:P05685
 C:Superfamily: chorion class A protein pc292
 F:1-4/Domain: signal sequence (fragment) #status predicted <SIG>
 F:5-151/Product: chorion class B protein #status predicted <MAT>

Query Match 35.8%; Score 105; DB 2; Length 151;
 Best Local Similarity 52.0%; Pred. No. 0.00092;
 Matches 26; Conservative 3; Mismatches 11; Indels 10; Gaps 4;

OY 4 GYGTSLLG---GSGVGP-YGSGFGS-YGSGVGYG-----GYGYGGY 43

Db 93 GIGBILYCGCGAVGITREGGLYGAGYGGYGLGYGGYGGY 142

RESULT 15
 B34504
 heterogeneous nuclear ribonucleoprotein B1 - human
 N:Alternate names: heterogeneous ribonuclear particle protein B1; hnRNP protein B1; NEPH
 N:Contains: heterogeneous nuclear ribonucleoprotein A2
 C:Species: Homo sapiens (man)
 C>Date: 22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change 09-Jul-2004
 A:Accession: A56845; B34504; A34504; S48057; PC2222; E61013; B56845
 R:Kozu, T.; Henrich, B.; Schaefer, K.P.
 Genomics 25, 365-371, 1995
 A:Title: Structure and expression of the gene (HNRPA2B1) encoding the human hnRNP protein
 A:Reference number: A56845; MUID:95309902; PMID:7789969
 A:Accession: A56845
 A>Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-353 <KOZ>
 A:Cross-references: UNIPROT:P22626; GB:D28877; NID:9565642; PIDN:BAA06031.1; PID:9565643
 R:Burd, C.G.; Swanson, M.S.; Goerlach, M.; Dreyfuss, G.
 Proc. Natl. Acad. Sci. U.S.A. 86, 9788-9792, 1989
 A:Title: Primary structures of the heterogeneous nuclear ribonucleoprotein A2, B1, and C
 A:Reference number: A34504; MUID:90099350; PMID:2557628
 A:Accession: B34504
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-353 <BUR>
 A:Cross-references: GB:M29064; NID:G337452; PIDN:AAA60271.1; PID:G337453
 A:Accession: A34504
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2,15-353 <BU2>
 A:Cross-references: GB:M29064
 R:Biamonti, G.; Ruggiu, M.; Saccone, S.; della Valle, G.; Riva, S.
 Nucleic Acids Res. 22, 1996-2002, 1994
 A:Title: Two homologous genes, originated by duplication, encode the human hnRNP protein
 A:Reference number: S48057; MUID:94301779; PMID:8029005
 A:Accession: S48057
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-353 <BIA>
 A:Cross-references: EMBL:U09122
 R:Prasad, S.; Walent, J.; Dritschilo, A.
 Biochem. Biophys. Res. Commun. 204, 772-779, 1994
 A:Title: ADP-ribosylation of heterogeneous ribonucleoproteins in HeLa cells.
 A:Reference number: PC2221; MUID:95071393; PMID:7980541
 A:Accession: PC2222
 A:Molecule type: protein
 A:Residues: 'XXX',83-86,'X',88-93,'X',95-100 <PRA>
 A:Experimental source: HeLa cells
 A:Note: this protein was shown to be ADP-ribosylated
 R:Bauw, G.; Rasmussen, H.H.; Van Den Bulcke, M.; Van Damme, J.; Puype, M.; Gesser, B.; C
 Electrophoresis 11, 528-536, 1990
 A:Title: Two-dimensional gel electrophoresis, protein electrophoresis and microsequencing
 A:Reference number: A61002; MUID:91031404; PMID:1699755
 A:Accession: E61013
 A:Molecule type: protein
 A:Residues: 63-69;204-212;214-220,'F',221-228 <BAU>
 C:Genetics:
 A:Gene: GDB:HNRPA2B1
 A:Cross-references: GDB:377778; OMIM:600124
 A:Map position: 7p15-7p15
 A:Introns: 2/3; 14/3; 51/3; 100/3; 171/1; 205/1; 232/1; 253/1; 293/1; 334/1
 C:Superfamily: helix-destabilizing protein; ribonucleoprotein repeat
 C:Keywords: alternative splicing; duplication; nucleus; RNA binding
 F:1-2,15-353/Domain: heterogeneous ribonuclear particle protein A2 <HA2>
 F:22-88/Domain: ribonucleoprotein repeat homology <RRM1>
 F:113-179/Domain: ribonucleoprotein repeat homology <RRM2>

Query Match 35.8%; Score 105; DB 2; Length 353;
 Best Local Similarity 50.0%; Pred. No. 0.002;
 Matches 23; Conservative 2; Mismatches 13; Indels 8; Gaps 2;

Search completed: April 1, 2005, 10:16:10
Job time : 5.32995 secs

BEST AVAILABLE COPY

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2005, 09:16:14 ; Search time 27.1827 Seconds
(without alignments)
725.636 Million cell updates/sec

Title: US-09-891-064A-2_COPY_196_246
Perfect score: 275
Sequence: 1 GWNPTAQSGLYSQIYAL.....LYDQYLHYCVDPQEAIA 51

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	275	100.0	508	8	Abm84985 Human dia
2	275	100.0	508	8	Abm84986 Human dia
3	275	100.0	522	2	Aaw36052 Human occ
4	275	100.0	522	2	Aaw34638 Human occ
5	275	100.0	522	3	Aab35731 Human occ
6	275	100.0	522	6	Abj37076 Human bre
7	275	100.0	522	7	Add46545 Human pro
8	275	100.0	522	8	Adi47189 Human occ
9	266	96.7	475	5	Aaul1278 Human rec
10	263	95.6	48	2	Aay17159 Peptide S
11	263	95.6	48	3	Aay96005 Human occ
12	263	95.6	48	3	Aay78216 Occludin
13	263	95.6	48	3	Aay64642 Occludin
14	263	95.6	48	4	Aag65438 Occludin
15	263	95.6	48	6	Abu60314 Mammalian
16	263	95.6	48	7	Abc43607 Classical
17	263	95.6	48	7	Abw01277 Occludin
18	263	95.6	48	8	Adk13612 Cadherin-
19	263	95.6	48	8	Ado70598 Occludin
20	263	95.6	48	8	Adp07175 Cell adhe
21	263	95.6	48	8	Adi74398 Human occ
22	263	95.6	594	5	Aaul1279 Human rec
23	255	92.7	521	2	Aaw36054 Mouse occ
24	251	91.3	47	2	Aay09102 Cell adhe
25	251	91.3	523	7	Add46543 Rat Prote

26	244	88.7	505	2	AAW36053	Canine oc
27	243	88.4	48	3	AAY96006	Mouse occ
28	243	88.4	48	8	ADR74399	Mouse occ
29	232	84.4	48	3	AAY96007	Dog occlu
30	232	84.4	48	8	ADR74400	Dog occlu
31	221.5	80.5	51	3	AAY96009	Mammalian
32	152	55.3	50	3	AAY96008	Rat kang
33	152	55.3	50	8	ADR74401	Rat-kang
34	133	48.4	44	3	AAB35733	Human occ
35	108	39.3	19	8	ADI46332	Permeabil
36	107	38.9	19	8	ADI46338	Permeabil
37	100	36.4	22	6	ABP70563	Peptide d
38	99	36.0	17	8	ADI46337	Permeabil
39	96	34.9	17	8	ADI46331	Permeabil
40	94	34.2	21	6	ABP70555	Analogue
41	94	34.2	21	6	ABP70559	Analogue
42	90	32.7	20	6	ABP70554	Analogue
43	90	32.7	20	6	ABP70560	Analogue
44	88	32.0	15	8	ADI46336	Permeabil
45	85	30.9	19	6	ABP70553	Analogue

ALIGNMENTS

RESULT 1
ABM84985
ID ABM84985 standard; protein; 508 AA.
XX
AC ABM84985;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human diagnostic and therapeutic pprotein SEQ ID NO:5234.
XX
KW gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX
OS Homo sapiens.
XX
PN WO2004023973-A2.
XX
PD 25-MAR-2004.
XX
PF 12-SEP-2003; 2003WO-US028227.
XX
PR 12-SEP-2002; 2002US-0410259P.
PR 12-SEP-2002; 2002US-0410260P.
XX
(INCY-) INCYTE CORP.
PA Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
Harthehorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
Stevens K, Blanchard JL, Panzer S, Wang X, Au AP, Gerstin EH;
Peralta CE, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton BS;
Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
Patury S, Shi X, Suarez CJ;
WPI; 2004-329368/30.
N-PSDB; ACN43637.
XX
DR New diagnostic and therapeutic polynucleotides and polypeptides, useful
in diagnosing a condition, disease or disorder associated with human
molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
in gene mapping.
XX
PS Claim 27; Page; 190pp; English.
XX
CC The invention relates to novel diagnostic and therapeutic polynucleotides
selected from one of the 2722 sequences defined in the specification. A
polynucleotide of the invention may have a use in gene therapy. The human
diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be

CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
XX Sequence 508 AA;

Query Match 100.0%; Score 275; DB 8; Length 508;
Best Local Similarity 100.0%; Pred. No. 1.5e-27;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVNPTAQSGLSGSIYALCNQFYTPAATGLYVDQYLYHYCVVDPEAIA 51
DB 196 GVNPTAQSGLSGSIYALCNQFYTPAATGLYVDQYLYHYCVVDPEAIA 246

RESULT 2

ABM84986
ID ABM84986 standard; protein; 508 AA.

AC ABM84986;

DT 18-NOV-2004 (first entry)

DE Human diagnostic and therapeutic pprotein SEQ ID NO:5235.

KW gene therapy; human diagnostic and therapeutic polynucleotide; dithp.

OS Homo sapiens.

PN WO2004023973-A2.

PD 25-MAR-2004.

PF 12-SEP-2003; 2003WO-US028227.

PR 12-SEP-2002; 2002US-0410259P.

PR 12-SEP-2002; 2002US-0410260P.

PA (INCY-) INCYTE CORP.

PI Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirtson ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;

XX WPI; 2004-329368/30.

DR N-PSDB; ACN43638.

XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.

PS Claim 27; Page; 190pp; English.

XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated

CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
XX Sequence 508 AA;

Query Match 100.0%; Score 275; DB 8; Length 508;
Best Local Similarity 100.0%; Pred. No. 1.5e-27;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVNPTAQSGLSGSIYALCNQFYTPAATGLYVDQYLYHYCVVDPEAIA 51
DB 196 GVNPTAQSGLSGSIYALCNQFYTPAATGLYVDQYLYHYCVVDPEAIA 246

RESULT 3

AAW36052
ID AAW36052 standard; protein; 522 AA.

AC AAW36052;

DT 25-MAR-1998 (first entry)

DE Human occludin protein sequence.

KW Occludin; adhesion molecule; membrane tight junction;
KW occludin localisation; membrane; occludin expression;
KW blood-brain barrier disorder.

OS Homo sapiens.

PN WO9732982-A1.

PD 12-SEP-1997.

PF 05-MAR-1997; 97WO-JP000665.

PR 07-MAR-1996; 96JP-00049880.

PR 12-DEC-1996; 96JP-00331944.

PA (EISA) EISAI CO LTD.

XX Tsukita S;

XX WPI; 1997-470546/43.

DR N-PSDB; AAT97972.

XX DNA encoding human, dog and mouse occludin(s) - useful for screening for
PT substances influencing occluding expression.

PS Claim 1; Page 15-17; 36pp; Japanese.

XX The present sequence represents human occludin, a novel adhesion molecule
CC which is a constituent protein of membrane tight junctions. Antibodies
CC raised against the protein are used to assay for occludin in samples, and
CC for the study of occludin localisation in membranes (e.g. by
CC immunofluorescent cell-staining). The transfectants and antibodies are
CC used for screening of substances which potentially influence occludin
CC expression. Therapeutic polynucleotides derived from the DNA encoding the
CC present sequence are used for treatment of disorders involving the blood-
CC brain barrier. Primers that hybridise to the occludin DNA sequence can be
CC used to detect occludin DNA by PCR

XX Sequence 522 AA;

Query Match 100.0%; Score 275; DB 2; Length 522;
Best Local Similarity 100.0%; Pred. No. 1.5e-27;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVNPTAQSSGLSGSIQYALCNQFYTPAATGLYVDQYLYHYCVVDPOQAIA 51
DB 196 GVNPTAQSSGLSGSIQYALCNQFYTPAATGLYVDQYLYHYCVVDPOQAIA 246

RESULT 4

AAW34638
ID AAW34638 standard; protein; 522 AA.

XX AC AAW34638;

XX DT 02-MAR-1998 (first entry)

XX DE Human occludin.

XX KW Occludin; inhibitor; human; drug absorption; drug delivery; gene therapy;
XX KW tissue permeability; cell adhesion.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide 90..138

FT FT /label= Extracellular loop

FT FT /note= "inhibitor peptide (Claim 9)"

FT FT Peptide 196..246

FT FT /label= Extracellular loop

XX XX /note= "inhibitor peptide (Claim 10)"

XX PN WO9733605-A1.

XX XX

XX PD 18-SEP-1997.

XX XX

XX PF 14-MAR-1997; 97WO-US005809.

XX PR 15-MAR-1996; 96US-0013625P.

XX PA (UYVA) UNIV YALE.

XX PI Anderson JM, Van Itallie CM;

XX XX WPI; 1997-470640/43.

XX DR N-PSDB; AAT93752.

XX XX

PT Isolated human occludin protein - useful for increasing drug delivery

PT across endothelial or epithelial barriers, or for reducing tissue

PT permeability.

XX Claim 2; Page 31-32; 49pp; English.

XX XX

CC This protein comprises human occludin, an integral transmembrane protein

CC specifically associated with tight junctions that functions in forming

CC intercellular seals. Its amino acid sequence was deduced from overlapping

CC cDNA clones (see AAT93752) obtained from a human liver cDNA library.

CC Occludin can be used to reduce tissue permeability. Occludin inhibitors

CC can be used to enhance delivery of drugs (or gene therapy vectors) by

CC increasing absorption across endothelial or epithelial barriers, i.e.

CC transmembrane or transvascular drug delivery. Inhibitors include occludin

CC surface loop peptides that inhibit adhesion and/or barrier properties, or

CC antibodies that interact with occludin or occludin receptors. Methods are

CC claimed for identifying occludin inhibitors and for screening for the

XX presence or absence of occludin inhibition in a sample

SQ Sequence 522 AA;

Query Match 100.0%; Score 275; DB 2; Length 522;

Best Local Similarity 100.0%; Pred. No. 1.5e-27;

Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVNPTAQSSGLSGSIQYALCNQFYTPAATGLYVDQYLYHYCVVDPOQAIA 51

DB 196 GVNPTAQSSGLSGSIQYALCNQFYTPAATGLYVDQYLYHYCVVDPOQAIA 246

RESULT 5

AAW35731
ID AAW35731 standard; protein; 522 AA.

XX AC AAW35731;

XX DT 19-FEB-2001 (first entry)

XX DE Human occludin amino acid sequence.

XX KW Pharmaceutical composition; gene therapy; junctional complex disruption;
XX KW cytoskeletal; cardiant; vasotropic; antiinflammatory; antilipemic;
XX KW cystic fibrosis; inflammatory disease; fibrotic lung disease;
XX KW peripheral vascular disease; coronary arterial disease; restenosis;
XX KW hypercholesterolaemia; cancer; human; occludin.

XX OS Homo sapiens.

XX PN WO200062815-A2.

XX PD 26-OCT-2000.

XX PF 13-APR-2000; 2000WO-GB001408.

XX PR 15-APR-1999; 99GB-00008636.

XX PR 15-APR-1999; 99GB-00008643.

XX PR 20-APR-1999; 99GB-00009073.

XX PR 22-APR-1999; 99GB-00009297.

XX PR 13-SEP-1999; 99US-0153757P.

XX PA (GLAX) GLAXO GROUP LTD.

XX PA (UYNC-) UNIV NORTH CAROLINA.

XX PI Boucher RC, Ford MJ, Johnson LG, Man Y, West MR;

XX DR WPI; 2000-679557/66.

XX XX

PT Novel pharmaceutical composition useful for gene therapy, comprises

PT nucleic acid encoding therapeutic gene and a promoter, transfection

PT vehicle and an agent which disrupts function of junctional complex in

PT cells.

XX Disclosure; Fig B; 51pp; English.

XX XX

CC This invention relates to a pharmaceutical composition used for gene

CC therapy, comprising a nucleic acid encoding a therapeutic gene and a

CC promoter, a transfection vehicle and an agent to disrupt the function of

CC the junctional complex in the cells. The pharmaceutical composition has

CC cytoskeletal, cardiant, vasotropic, antiinflammatory and antilipemic

CC activity. The pharmaceutical composition is useful for gene therapy and

CC useful for enhancing efficiency of gene delivery to cells. The composition is

CC useful for gene therapy of cystic fibrosis, inflammatory or fibrotic lung

CC diseases, peripheral vascular disease, coronary arterial diseases,

CC restenosis, hypercholesterolaemia and cancer. The present sequence

CC represents the human occludin protein. Occludin is used in the invention

CC to determine the effect of the junctional inhibitors used in the

XX pharmaceutical complex

SQ Sequence 522 AA;

Query Match 100.0%; Score 275; DB 3; Length 522;

Best Local Similarity 100.0%; Pred. No. 1.5e-27;

Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVNPTAQSSGLSGSIQYALCNQFYTPAATGLYVDQYLYHYCVVDPOQAIA 51

DB 196 GVNPTAQSSGLSGSIQYALCNQFYTPAATGLYVDQYLYHYCVVDPOQAIA 246

```
RESULT 6
ABJ37076
ID ABJ37076 standard; protein; 522 AA.
XX
AC ABJ37076;
XX
DT 01-MAY-2003 (first entry)
XX
DE Human breast cancer / ovarian cancer related protein #52.
XX
KW Human; cytostatic; breast cancer; ovarian cancer.
XX
OS Homo sapiens.
XX
PN WO2003000012-A2.
XX
PD 03-JAN-2003.
XX
PF 21-JUN-2002; 2002WO-US019773.
XX
PR 21-JUN-2001; 2001US-0300159P.
PR 27-JUN-2001; 2001US-0301351P.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Velby OP;
XX
WPI; 2003-267848/26.
DR N-PSDB; ABT31945.
XX
PT Determining the presence of breast cancer in an individual, involves
PT using specific polynucleotide markers.
XX
PS Disclosure; Page 223-224; 233pp; English.
XX
CC The invention comprises a method for assessing whether a patient is
CC afflicted with breast cancer or ovarian cancer. The method involves the
CC use of specific DNA markers. The method of the invention is useful in the
CC detection and treatment of ovarian and breast cancer. Amino acid
CC sequences ABJ37025 - ABJ37080 represent human breast/ovarian cancer-
CC related proteins
XX
SQ Sequence 522 AA;
Query Match 100.0%; Score 275; DB 6; Length 522;
Best Local Similarity 100.0%; Pred. No. 1.5e-27;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GVNPTAQSGSLGSGIYALCNQFYTPAATGLYVDQYLYHYCVVDPQEAIA 51
DB 196 GVNPTAQSGSLGSGIYALCNQFYTPAATGLYVDQYLYHYCVVDPQEAIA 246
RESULT 7
ADD46545
ID ADD46545 standard; protein; 522 AA.
XX
AC ADD46545;
XX
DT 02-DEC-2004 (revised)
DT 29-JAN-2004 (first entry)
XX
DE Human Protein NP_002529, SEQ ID NO 12226.
XX
KW Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX
OS Homo sapiens.
OS Unidentified.
XX
PN WO2003016475-A2.
XX
```

```
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
PA (GEHO ) GEN HOSPITAL CORP.
PA (FARB ) BAYER AG.
XX
Woolf C, D'urso D, Befort K, Costigan M;
XX
WPI; 2003-268312/26.
DR GENBANK; NP_002529.
XX
New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
Example 1; Page; 1017pp; English.
XX
CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (described in Table 3
CC of the specification) which is differentially expressed during pain.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 522 AA;
Query Match 100.0%; Score 275; DB 7; Length 522;
Best Local Similarity 100.0%; Pred. No. 1.5e-27;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GVNPTAQSGSLGSGIYALCNQFYTPAATGLYVDQYLYHYCVVDPQEAIA 51
DB 196 GVNPTAQSGSLGSGIYALCNQFYTPAATGLYVDQYLYHYCVVDPQEAIA 246
RESULT 8
ADI47189
ID ADI47189 standard; protein; 522 AA.
XX
AC ADI47189;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human occludin protein sequence.
XX
KW mucosal delivery; permeabilising peptide;
KW mucosal epithelial paracellular transport; epithelial junction;
KW epithelial membrane adhesive protein; junctional adhesion molecule; JAM;
XX
```


KW occludin; claudin; sexual dysfunction; male erectile sexual dysfunction;
 KW female sexual dysfunction; human.
 XX
 OS Homo sapiens.
 XX WO2004003145-A2.
 PN
 XX
 XX 08-JAN-2004.
 PD
 XX 24-JUN-2003; 2003WO-US019994.
 XX
 XX 28-JUN-2002; 2002US-0392512P.
 PR
 XX (NAST-) NASTECH PHARM CO INC.
 PA
 XX Quay SC;
 PI
 XX WPI; 2004-091087/09.
 DR
 XX Composition comprising biologically active agent and permeabilizing
 PT peptide that reversibly enhances mucosal epithelial paracellular
 PT transport by modulating epithelial junctional structure and/or physiology
 PT in mammal.
 XX
 PS Disclosure; SEQ ID NO 900; 426pp; English.
 XX
 XX This invention relates to a novel composition comprising a biologically
 CC active agent and mucosal delivery-enhancing effective amount of
 CC permeabilizing peptide that reversibly enhances mucosal epithelial
 CC paracellular transport by modulating epithelial junctional structure
 CC and/or physiology in a mammal. The agent of the invention inhibits
 CC homotypic binding of epithelial membrane adhesive protein chosen
 CC functional adhesion molecule (JAM), occludin and claudin. The
 CC biologically active agent is effective for treatment of sexual
 CC dysfunction, for example male erectile sexual dysfunction or female
 CC sexual dysfunction. The present sequence is that of the human occludin
 CC protein which is related to the invention.
 XX
 SQ Sequence 522 AA;

Query Match 100.0%; Score 275; DB 8; Length 522;
 Best Local Similarity 100.0%; Pred. No. 1.5e-27;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GVNPTAQSSGLSGSIYALCNQFYTPAATGLYVDQYLYHYCVVDPQBAIA 51
 DB 196 GVNPTAQSSGLSGSIYALCNQFYTPAATGLYVDQYLYHYCVVDPQBAIA 246

RESULT 9
 AAU11278
 ID AAU11278 standard; protein; 475 AA.
 XX
 AC AAU11278;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human recombinant J36 polypeptide.
 XX
 KW Human; KV protein; KVE702; angiogenic factor antagonist; angiogenesis;
 KW vascular endothelial structure regulator; cancer; vascular endothelium;
 KW tumour; blood vessel; epithelial cell; cytostatic; gene therapy; J36.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 48..52
 FT /note= "Encoded by GAGCCA"
 FT
 XX WO200183562-A2.
 PN
 XX 08-NOV-2001.

XX 04-MAY-2001; 2001WO-GB001956.
 PF
 XX
 PR 04-MAY-2000; 2000GB-00010630.
 XX
 PA (UTWA-) UNIV WALES COLLEGE OF MEDICINE.
 XX
 PI Jiang WG;
 XX
 DR WPI; 2002-066521/09.
 DR N-PSDB; AAS16973.
 XX
 PT Novel nucleic acid for inhibition of angiogenesis and prophylaxis or
 PT treatment of cancer, comprises both an angiogenic factor antagonist and a
 PT vascular endothelial structure regulator.
 XX
 PS Disclosure; Fig 14; 66pp; English.
 XX
 CC The invention relates to recombinant nucleic acid sequences that encode
 CC both an angiogenic factor antagonist and a vascular endothelial structure
 CC regulator, collectively referred to as a KV protein. KV DNA and proteins,
 CC and also cells and compositions containing the sequences, are useful in
 CC treating or preventing cancer or angiogenesis in mammals, in particular
 CC humans. KV proteins are also useful for regulation of the development of
 CC blood vessels and their formation, in the vascular endothelium and/or
 CC tumour. The KVE702 gene and its fragments are useful in transfection of
 CC human epithelial cells and to generate products suitable for angiogenesis
 CC intervention. This sequence represents the human J36 protein used in
 CC methods of the invention
 XX
 SQ Sequence 475 AA;
 Query Match 96.7%; Score 266; DB 5; Length 475;
 Best Local Similarity 98.0%; Pred. No. 2.1e-26;
 Matches 49; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GVNPTAQSSGLSGSIYALCNQFYTPAATGLYVDQYLYHYCVVDPQBAI 50
 DB 1 GVNPTAQSSGLSGSIYALCNQFYTPAATGLYVDQYLYHYCVVDPQBAI 50

RESULT 10
 AAU17159
 ID AAU17159 standard; peptide; 48 AA.
 XX
 AC AAU17159;
 XX
 DT 20-SEP-1999 (first entry)
 XX
 DE Peptide Seq ID No: 59 of WO9933875.
 XX
 KW Cell adhesion modulation; CAM; synaptic stability; cadherin;
 KW cadherin-mediated adhesion; drug delivery; cell adhesion; tumour;
 KW wound healing; neurite outgrowth.
 XX
 OS Synthetic.
 OS
 PN WO9933875-A1.
 XX
 PD 08-JUL-1999.
 XX
 PF 23-DEC-1998; 98WO-CA001207.
 XX
 PR 23-DEC-1997; 97US-00996679.
 XX
 PA (UYWC-) UNIV MCGILL.
 XX
 PI Blaschuk OW, Gour BJ;
 XX
 DR WPI; 1999-430231/36.
 XX
 PT Cyclic peptide cell adhesion modulating agents, useful for modulating
 PT synaptic stability.

XX Example; Page 138; 144pp; English.

PS The invention provides cyclic peptide cell adhesion modulating (CAM)

XX agents that comprises a His-Ala-Val recognition sequence. Also provided

CC is a method for inhibiting synaptic stability in a mammal that comprises

CC administering to a mammal a therapeutically effective amount of a CAM

CC agent that inhibits cadherin-mediated adhesion, where the agent comprises

CC a cyclic peptide having a peptide ring, and where the sequence His-Ala-

CC Val is present within the peptide ring. The cyclic peptides are cell

CC adhesion modulating agents that inhibit cadherin-mediated adhesion. They

CC can be used in a method for inhibiting synaptic stability in mammals. The

CC agents can be used to treat diseases or other conditions characterized by

CC undesirable cell adhesion or to facilitate drug delivery to a specific

CC tissue or tumour. Alternatively the agents may be used to enhance cell

CC adhesion (e.g. to supplement or replace stitches or to facilitate wound

CC healing) or to enhance or direct neurite outgrowth

XX

SQ Sequence 48 AA;

Query Match 95.6%; Score 263; DB 2; Length 48;

Best Local Similarity 100.0%; Pred. No. 3.5e-27;

Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVNPTAQSSGLYSGQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48

DB 1 GVNPTAQSSGLYSGQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48

RESULT 11

AAV96005

ID AAY96005 standard; peptide; 48 AA.

XX

AC AAY96005;

XX

DT 05-DEC-2000 (first entry)

XX

DE Human occludin extracellular domain 2.

XX

KW Cell adhesion; occludin; human; tissue permeability; vasopermeability;

KW vulnary; wound healing; drug delivery; tumour; therapy.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Region 38..41

FT /note= "cell adhesion recognition sequence"

XX

PN US6110747-A.

XX

PD 29-AUG-2000.

XX

PF 29-DEC-1998; 98US-00222373.

XX

PR 31-DEC-1997; 97US-00001511.

XX

PA (ADHE-) ADHEREX TECHNOLOGIES INC.

XX

PI Symonds JM, Gour BJ, Blaschuk OW;

XX

DR WPI; 2000-586358/55.

XX

PT Use of modulating agent comprising an occludin sequence to identify

PT compound that modulates occludin-mediated cell adhesion, involves

PT contacting test compound with antibody and detecting antibody binding

PT level.

XX

PS Disclosure; Fig 2; 48pp; English.

XX

CC The present sequence is that of the second extracellular domain of human

CC occludin. This domain of occludin includes a cell adhesion recognition

CC (CAR) motif that promotes the formation of permeability barriers. The CAR

CC motif is also found in other mammalian occludins (see AAY96006-8). Linear

CC (see AAY95990-Y95997) and cyclic (see AAY95998-Y96004) peptides of the

CC invention include the the CAR sequence. They are useful cell adhesion

CC modulating agents capable of modulating occludin-mediated processes, such

CC as cell adhesion. The invention provides methods for using modulating

CC agents to enhance or inhibit occludin-mediated cell adhesion in vivo or

CC in vitro. The modulating agents comprise at least 1 occludin CAR sequence

CC or an antibody that specifically binds the occludin CAR sequence. The

CC modulating agents may be linked to a targeting agent, drug and/or support

CC material. Agents that inhibit cell adhesion may be used to treat

CC conditions characterized by undesirable cell adhesion or to facilitate

CC (transdermal) drug delivery to a specific tissue or tumour. Agents that

CC enhance cell adhesion may be used to enhance wound healing or reduce

CC scarring. Antibodies raised against peptides containing the occludin CAR

CC sequence are used in a claimed method for identifying a compound capable

CC of modulating occludin mediated cell adhesion, especially when the test

CC compound is produced by an expression library

XX

SQ Sequence 48 AA;

Query Match 95.6%; Score 263; DB 3; Length 48;

Best Local Similarity 100.0%; Pred. No. 3.5e-27;

Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVNPTAQSSGLYSGQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48

DB 1 GVNPTAQSSGLYSGQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48

RESULT 12

AAV78216

ID AAY78216 standard; peptide; 48 AA.

XX

AC AAY78216;

XX

DT 03-MAY-2000 (first entry)

XX

DE Occludin cell adhesion recognition sequence SEQ ID NO:69.

XX

KW Cadherin; modulation; cell adhesion recognition; CAR; HAV-BF;

KW HAV binding motif; antiapoptotic; cytostatic; cell adhesion;

KW neurite outgrowth; Schwann cell migration; tumour; cancer; metastasis;

KW apoptosis; angiogenesis; immune system; pregnancy; vasopermeability;

KW demyelinating neurological disease.

XX

OS Unidentified.

XX

PN WO200002917-A2.

XX

PD 20-JAN-2000.

XX

PF 12-JUL-1999; 99WO-CA000627.

XX

PR 10-JUL-1998; 98US-00113977.

XX

PA (ADHE-) ADHEREX TECHNOLOGIES INC.

XX

PI Doherty P, Blaschuk OW, Gour BJ;

XX

DR WPI; 2000-160899/14.

XX

PT Modulating agents for treating autoimmune diseases, cancer, spinal cord

PT injuries, and for increasing vasopermeability, inhibiting synaptic

PT stability and preventing pregnancy.

XX

PS Disclosure; Page 22; 144pp; English.

XX

CC The present invention describes a cell adhesion modulating agent (I)

CC capable of binding to the cadherin cell adhesion recognition (CAR)

CC sequence HAV, where the agent doesn't comprise an antibody or antigen-

CC binding fragment of it. (i) is used for inhibiting or enhancing cadherin

CC mediated functions like cell adhesion, neurite outgrowth, Schwann cell

CC migration and synaptic stability in cells preferably epithelial,

CC endothelial, neural, tumour cells and lymphocytes expressing cadherin E

or N. Inhibition of cadherin mediated cell adhesion by (I) is used in reducing unwanted cellular adhesion, enhancing drug delivery through skin, drug delivery to a tumour, treating cancer and/or inhibiting metastasis, inducing apoptosis, inhibiting angiogenesis, modulating immune system, preventing pregnancy, increasing vasopermeability, inhibiting synaptic stability in a mammal. Enhancement is used for facilitating wound healing, enhancing adhesion of foreign tissue implant, enhancing and/or directing neurite outgrowth and treating spinal cord injuries. (I) is used for treating demyelinating neurological diseases and for facilitating blood sampling in a mammal. (I) is used for detecting cadherin expressing cells in a sample by detecting the level of antibody cadherin complex. AAY78148 to AAY78242 represent sequences used in the exemplification of the present invention

XX Sequence 48 AA;
SQ
Query Match 95.6%; Score 263; DB 3; Length 48;
Best Local Similarity 100.0%; Pred. No. 3.5e-27;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVNPTAQSSGLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVDPQ 48
DB 1 GVNPTAQSSGLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVDPQ 48

RESULT 13
AAY64642
ID AAY64642 standard; peptide; 48 AA.
XX
AC AAY64642;
XX

DT 02-MAR-2000 (first entry)

XX Occludin cell adhesion recognition sequence SEQ ID NO:62.

DE Modulation; nonclassical cadherin mediated cell adhesion; CAR;
XX inhibition; cadherin extracellular domain; cell adhesion recognition;
KW OB-cadherin; cadherin-5; cadherin-6; cadherin-7; cadherin-8; cadherin-12;
KW cadherin-14; cadherin-15; T-cadherin; PB-cadherin;
KW cadherin related neuronal receptor; Li-cadherin; protocadherin;
KW desmoglein; desmocollin; calcium binding; cancer; tumour; obesity;
KW rheumatoid arthritis; multiple sclerosis; diabetes; metastasis;
KW neurological disease.

XX Synthetic.

OS Homo sapiens.

XX WO957149-A2.

XX

PN 11-NOV-1999.

XX

PD

XX

PF 05-MAY-1999; 99WO-CA000363.

XX

PR 05-MAY-1998; 98US-00073040.

XX

PR 06-NOV-1998; 98US-00187859.

XX

PR 20-JAN-1999; 99US-00234395.

XX

PR 08-MAR-1999; 99US-00264516.

XX

XX

PA (ADHE-) ADHEREX TECHNOLOGIES INC.

XX

PI Blaschuk OW, Gour BJ, Byers S;

XX

XX WPI; 2000-038791/03.

XX

DR

XX New cadherin modulating agents, used for modulating nonclassical cadherin

PT mediated functions for treating e.g. cancers, obesity, rheumatoid

PT arthritis, multiple sclerosis, diabetes or a neurological disease.

XX

XX Disclosure; Page 133; 252pp; English.

CC nonclassical cadherin-mediated functions. They can be used for e.g. inhibiting adhesion of nonclassical-cadherin expressing cells in a mammal, enhancing delivery of a drug through the skin of a mammal, enhancing delivery of a drug to a tumour in a mammal, treating cancer in a mammal, inhibiting metastasis of a cancer in a mammal, inhibiting angiogenesis in a mammal, inducing apoptosis in a nonclassical cadherin-expressing cell, preventing or treating obesity in a mammal, stimulating blood vessel regression in a mammal, enhancing drug delivery to the central nervous system, treating a demyelinating neurological disease, increasing vasopermeability in a mammal, enhancing adhesion of nonclassical cadherin-expressing cells, inhibiting synaptic stability in a mammal, or preventing pregnancy in a mammal. They can also be used for e.g. enhancing or directing neurite outgrowth, facilitating wound healing or reducing scar tissue, or enhancing adhesion of foreign tissue in a mammal. They can also be used for treating e.g. psoriasis, arthritis, age-related macular degeneration, multiple sclerosis and diabetes. The products can also be used for detection and diagnosis and in bioreactors. CC AAY60592 to AAY64572 represent specifically claimed peptides, and CC AAY64573 to AAY64643 and AAZ33183 to AAZ33186 represent sequences used in the exemplification of the present invention

XX Sequence 48 AA;

SQ
Query Match 95.6%; Score 263; DB 3; Length 48;
Best Local Similarity 100.0%; Pred. No. 3.5e-27;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVNPTAQSSGLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVDPQ 48

DB 1 GVNPTAQSSGLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVDPQ 48

RESULT 14

AAG65438

ID AAG65438 standard; peptide; 48 AA.

XX

AC AAG65438;

XX

DT 30-NOV-2001 (first entry)

XX

DE Occludin cell adhesion recognition sequence.

XX Cell adhesion; imidazole; peptidomimetic; cadherin; angiogenesis; CAR;

KW cytostatic; vulnary; immunomodulator; vasotropic; neuroprotective;

KW cerebroprotective; muscular; occludin.

XX Unidentified.

OS

XX WO200153331-A2.

XX

PN 26-JUL-2001.

XX

PD

XX

PF 24-JAN-2001; 2001WO-US002508.

XX

PR 24-JAN-2000; 2000US-00491078.

XX

XX (ADHE-) ADHEREX TECHNOLOGIES INC.

XX

PI Gour BJ, Blaschuk OW, Ali A, Ni F, Chen Z, Michaud SD, Wang S;

XX Hu Z;

XX WPI; 2001-549899/61.

XX

DR Cell adhesion modulating agent used for enhancing delivery of drug to

PT tumor comprises imidazole compounds.

XX

XX Example; Page 410; 436pp; English.

XX

CC The invention relates to cell adhesion modulating agents that comprise

CC imidazole compounds of specified formulae that are peptidomimetics of

CC cyclic peptides. The peptidomimetics have a structure similar to that of

CC a cyclic peptide that comprises a cadherin cell recognition sequence HAV.

CC The agents are used for modulating classical cadherin mediated

CC intercellular adhesion, reducing unwanted cellular adhesion, enhancing
 CC delivery of a drug to a tumor, inhibiting development of cancer,
 CC inhibiting angiogenesis, enhancing drug delivery to the CNS, enhancing
 CC wound healing, modulating the immune system, increasing vasopermeability,
 CC treating demyelinating diseases, facilitating migration of an N-cadherin
 CC expressing cell on astrocytes, inhibiting synaptic stability, modulating
 CC neurite outgrowth, and treating spinal cord injuries and macular
 CC degeneration. The present sequence represents an occludin cell adhesion
 CC recognition (CAR) sequence with flanking amino acids
 XX

SQ Sequence 48 AA;

Query Match 95.6%; Score 263; DB 4; Length 48;
 Best Local Similarity 100.0%; Pred. No. 3.5e-27;
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVNPTAQSSGLYGSQIYALCNQFYTPAATGLYVDQYLHYCVDPQE 48

Db 1 GVNPTAQSSGLYGSQIYALCNQFYTPAATGLYVDQYLHYCVDPQE 48

RESULT 15

ABU60314

ID ABU60314 standard; peptide; 48 AA.

XX

AC ABU60314;

XX

DT 29-APR-2003 (first entry)

XX

DE Mammalian occludin CAR sequence #1.

XX

KW Cell adhesion; cadherin-mediated function; His-Ala-Val binding motif;
 KW HAV-BM; cell adhesion recognition; CAR; adhesion molecule; tumour;
 KW neurite outgrowth; Schwann cell migration; synaptic stability; T-cell;
 KW drug delivery; central nervous system; CNS; cancer; metastasis;
 KW apoptosis; angiogenesis; wound healing; foreign tissue; implantation;
 KW spinal cord injury; demyelinating neurological disease; immune system;
 KW multiple sclerosis; pregnancy; vasopermeability; blood sampling;
 KW diabetes; rheumatoid arthritis; skin disorder; cytostatic; vulnary;
 KW cutaneous lymphoma; antirheumatic; antiarthritic; neuroprotective;
 KW antidiabetic; contraceptive; mammalian; occludin CAR.

XX Mammalia.

XX

PN US6472368-B1.

XX

PD 29-OCT-2002.

XX

PF 09-JUL-1999; 99US-00351048.

XX

PR 10-JUL-1999; 98US-00113977.

XX

PA (ADHE-) ADHEREX TECHNOLOGIES INC.

XX

PI Doherty P, Blaschuk OW, Gour BJ;

XX

DR WPI; 2003-265085/26.

XX

PT Cell adhesion modulating agent, useful for modulating cadherin-mediated
 PT functions, e.g. cell adhesion, neurite outgrowth, Schwann cell migration
 PT and synaptic stability, consists of a cyclic peptide.

XX

PS Disclosure; Col 37; 63pp; English.

XX

CC The present invention relates to cell adhesion modulating agents and
 CC methods for enhancing or inhibiting cadherin-mediated functions. The cell
 CC adhesion modulating agents comprise at least a His-Ala-Val binding motif
 CC (HAV-BM), an analogue or peptidomimetic of, or an antibody or fragment
 CC that specifically binds to such a motif. The modulating agents may also
 CC comprise one or more cell adhesion recognition sequences (CAR) recognised
 CC by cadherin and/or other adhesion molecules. Preferably the modulating
 CC agents are cyclic peptides. The modulating agents of the invention are
 CC useful for enhancing or inhibiting cadherin-mediated functions (e.g. cell

CC adhesion, neurite outgrowth, Schwann cell migration and synaptic
 CC stability), enhancing the delivery of a drug through the skin of a mammal
 CC or to a tumour or central nervous system (CNS) of a mammal, treating
 CC cancer and/or inhibiting metastases in a mammal, inducing apoptosis in a
 CC cadherin-expressing cell, inhibiting angiogenesis in a mammal,
 CC facilitating wound healing in a mammal, enhancing adhesion of foreign
 CC tissue implanted within a mammal, enhancing and/or directing neurite
 CC outgrowth, treating spinal cord injuries, treating demyelinating
 CC neurological diseases (e.g. multiple sclerosis), modulating the immune
 CC system, preventing pregnancy, increasing vasopermeability, controlling
 CC inhibition of synaptic stability, facilitating blood sampling, treating
 CC diseases associated with excess generation of otherwise normal T cells,
 CC treating diabetes, rheumatoid arthritis, and skin disorders (e.g.
 CC cutaneous lymphomas). The present sequence represents a mammalian
 CC occludin CAR sequence
 XX

SQ Sequence 48 AA;

Query Match 95.6%; Score 263; DB 6; Length 48;

Best Local Similarity 100.0%; Pred. No. 3.5e-27;

Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVNPTAQSSGLYGSQIYALCNQFYTPAATGLYVDQYLHYCVDPQE 48

Db 1 GVNPTAQSSGLYGSQIYALCNQFYTPAATGLYVDQYLHYCVDPQE 48

Search completed: April 1, 2005, 10:10:10

Job time : 28.1827 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2005, 09:31:32 ; Search time 24.7665 Seconds
(without alignments)
1054.491 Million cell updates/sec

Title: US-09-891-064a-2_COPY_196_246

Perfect score: 275

Sequence: 1 GWNPTAQSGSLYSGIYAL.....LYVDQYLHYVCVDPQEAIA 51

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	275	100.0	522	1 OCLN HUMAN	Q16625 homo sapien
2	266	96.7	449	2 Q6PP84	Q6PP84 bos taurus
3	255	92.7	521	1 OCLN MOUSE	Q61146 mus musculus
4	251	91.3	523	2 Q6P6T5	Q6P6T5 rattus norv
5	251	91.3	523	2 Q9Z303	Q9Z303 rattus norv
6	244	88.7	521	1 OCLN CANFA	Q28269 canis fami
7	244	88.7	560	2 Q9N0W3	Q9N0W3 canis fami
8	204	74.2	493	1 OCLN XENLA	Q9PUN1 xenopus lae
9	204	74.2	519	2 Q6GJ48	Q6GJ48 xenopus lae
10	167.5	60.9	497	2 Q6NRR0	Q6NRR0 xenopus lae
11	165	60.0	492	2 Q7ZUE7	Q7ZUE7 brachydanio
12	163	59.3	489	1 OCLN POTTR	Q28793 potorous tr
13	162	58.9	500	2 Q6NX99	Q6NX99 brachydanio
14	144	52.4	504	1 OCLN CHICK	Q91049 gallus gall
15	76	27.6	81	2 Q91XW5	Q91XW5 rattus norv
16	66	24.0	406	2 Q8VDS1	Q8VDS1 mus musculus
17	66	24.0	1246	2 Q9NEE0	Q9NEE0 leishmania
18	65.5	23.8	225	2 Q8LKK9	Q8LKK9 phytophthor
19	65	23.6	1085	2 Q7TP51	Q7TP51 rattus norv
20	64.5	23.5	1386	2 Q91TL1	Q91TL1 tupaiid her
21	63.5	23.1	236	2 Q8LKL0	Q8LKL0 phytophthor
22	63.5	23.1	238	2 Q9AT04	Q9AT04 phytophthor
23	61	22.2	505	2 Q6FL57	Q6FL57 candida gla
24	60.5	22.0	237	2 Q9AT28	Q9AT28 phytophthor
25	60	21.8	182	2 Q8G9X5	Q8G9X5 escherichia
26	60	21.8	533	2 Q9A5H0	Q9A5H0 caulobacter
27	60	21.8	691	2 Q7PKC4	Q7PKC4 anopheles g
28	60	21.8	853	2 Q6UE41	Q6UE41 humulus jap
29	59	21.5	180	2 Q96T87	Q96T87 homo sapien
30	59	21.5	198	2 Q8XWD4	Q8XWD4 raietonia s
31	59	21.5	544	2 Q6FT62	Q6FT62 candida gla

32 58.5 21.3 237 2 Q9AXL8
33 58 21.1 100 1 Q921 RHOMO
34 58 21.1 294 2 Q926Y3
35 58 21.1 462 1 COAT IRV16
36 58 21.1 554 2 Q6LVR8
37 58 21.1 956 2 Q6N36
38 57.5 20.9 199 2 Q7R4D6
39 57.5 20.9 327 2 Q9YU11
40 57.5 20.9 337 2 Q8EM97
41 57.5 20.9 543 2 Q8GA61
42 57.5 20.9 553 2 Q9X1V5
43 57.5 20.9 933 2 Q8IMR1
44 57.5 20.9 996 2 Q8IMR0
45 57.5 20.9 1343 2 Q7TFL9

ALIGNMENTS

RESULT 1

ID OCLN HUMAN STANDARD; PRT; 522 AA.
AC Q16625; Q9N6K1;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Occludin.
GN Name=OCLN;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon carcinoma;
RX MEDLINE=96181088; PubMed=8601611; DOI=10.1083/jcb.133.1.43;
RA Ando-Akatsuka Y., Saitou M., Hirase T., Kishi M., Sakakibara A.,
Itoh M., Yonemura S., Furuse M., Tsukita S.;
RT "Interspecies diversity of the occludin sequence: cDNA cloning of
human, mouse, dog, and rat-kangaroo homologues.";
RL J. Cell Biol. 133:43-47(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA van Itallie C.M., Fanning A.S., Anderson J.M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Fukasawa M., Toyota T., Yoshitsugu K., Yoshikawa T.;
RL "Genomic structure of occludin gene.";
RN Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, and Lung;
RX MEDLINE=23388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Mox S.I., Wang J., Hsieh F.,
Diatchenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.":
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: May play a role in the formation and regulation of the
 CC tight junction (TJ) paracellular permeability barrier.
 CC -!- SUBUNIT: Interacts with TJP1/ZO1 and with VAPA.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: Localized at tight junctions of both
 CC epithelial and endothelial cells. Highly expressed in kidney. Not
 CC detected in testis.
 CC -!- DOMAIN: The C-terminal is cytoplasmic and is important for
 CC interaction with ZO-1. Sufficient for the tight junction
 CC localization. Involved in the regulation of the permeability
 CC barrier function of the tight junction (By similarity).
 CC -!- PTM: Phosphorylated (By similarity).
 CC -!- SIMILARITY: Belongs to the ELL / occludin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U49184; AAC50451.1; -;
 CC EMBL; U53823; AAB00195.1; -;
 CC EMBL; AF400630; AAL47094.1; -;
 CC EMBL; AF400623; AAL47094.1; JOINED.
 CC EMBL; AF400624; AAL47094.1; JOINED.
 CC EMBL; AF400625; AAL47094.1; JOINED.
 CC EMBL; AF400626; AAL47094.1; JOINED.
 CC EMBL; AF400627; AAL47094.1; JOINED.
 CC EMBL; AF400628; AAL47094.1; JOINED.
 CC EMBL; AF400629; AAL47094.1; JOINED.
 CC EMBL; BC029886; AAL47094.1; JOINED.
 CC PIR; G02533; G02533.
 CC Genew; HGNC:8104; OCLN.
 CC H-InvDB; HX0019160; -;
 CC MIM; 602876; -;
 CC GO; GO:0005886; C:plasma membrane; TAS.
 CC GO; GO:0006461; P:protein complex assembly; TAS.
 CC InterPro; IPR008253; Marvel.
 CC InterPro; IPR002958; Occludin.
 CC InterPro; IPR010844; Occludin_ELL.
 CC Pfam; PF01284; MARVEL; 1.
 CC Pfam; PF07303; Occludin_ELL; 1.
 CC PRINTS; PR01258; OCLUDIN.
 CC Coiled coil; Phosphorylation; Tight junction; Transmembrane.
 CC DOMAIN 1 66 Cytoplasmic (Potential).
 CC TRANSMEM 67 89 Potential.
 CC DOMAIN 90 135 Extracellular (Potential).
 CC TRANSMEM 136 160 Potential.
 CC DOMAIN 161 170 Cytoplasmic (Potential).
 CC TRANSMEM 171 195 Potential.
 CC DOMAIN 196 243 Extracellular (Potential).
 CC TRANSMEM 244 265 Potential.
 CC DOMAIN 266 522 Cytoplasmic (Potential).
 CC DOMAIN 92 131 Gly/tyr-rich.
 CC DOMAIN 426 489 Coiled coil (Potential).
 CC CONFLICT 233 233 L -> S (in Ref. 4).
 CC SEQUENCE 522 AA; 59143 MW; AOCF9574BCF6E974 CRC64;

Query Match 100.0%; Score 275; DB 1; Length 522;
 Best Local Similarity 100.0%; Pred. No. 4.4e-27;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVNPTAQSSGSLYSQIYALCNQFYTPATGLYVDQYLHYCVVDPOEAIA 51
 DB 196 GVNPTAQSSGSLYSQIYALCNQFYTPATGLYVDQYLHYCVVDPOEAIA 246

RESULT 2
 Q6PP84

ID Q6PP84 PRELIMINARY; PRT; 449 AA.
 AC Q6PP84;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Occludin (Fragment).
 GN Name=OCLN;
 OS Bos taurus (Bovine).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC Bovinae; Bos.
 CC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=retinal endothelium;
 RA Sundstrom J.M., Felinski E.A., Phillips B., Antonetti D.A.;
 RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY589500; AAT00455.1; -;
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0005923; C:tight junction; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR008253; Marvel.
 DR InterPro; IPR002958; Occludin.
 DR InterPro; IPR010844; Occludin_ELL.
 DR Pfam; PF01284; MARVEL; 1.
 DR Pfam; PF07303; Occludin_ELL; 1.
 DR PRINTS; PR01258; OCLUDIN.
 FT NON TER 1
 FT SEQUENCE 449 AA; 50627 MW; 235B8259CF05D1C CRC64;
 SQ
 Query Match 96.7%; Score 266; DB 2; Length 449;
 Best Local Similarity 96.1%; Pred. No. 5.6e-26;
 Matches 49; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GVNPTAQSSGSLYSQIYALCNQFYTPATGLYVDQYLHYCVVDPOEAIA 51
 DB 123 GVNPTAQSSGSLYSQIYALCNQFYTPATGLYVDQYLHYCVVDPOEAIA 173
 RESULT 3
 ID OCLN_MOUSE STANDARD; PRT; 521 AA.
 AC Q61146;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE Occludin.
 GN Name=Ocln; Synonyms=Ocl;
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=96181088; PubMed=8601611; DOI=10.1083/jcb.133.1.43;
 RA Ando-Akatsuka Y., Saitou M., Hirase T., Kishi M., Sakakibara A.,
 RA Itoh M., Yonemura S., Furuse M., Tsukita S.;
 RA "Interspecies diversity of the occludin sequence: cDNA cloning of
 RA human, mouse, dog, and rat-kangaroo homologues."; J. Cell Biol. 133:43-47(1996).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=C57BL/6J; TISSUE=Ovary, and Uterus;
 RC MEDLINE=22354683; PubMed=12468851; DOI=10.1038/nature01266;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Mikado I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
 RA Nagai K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bradt D., Brusich V., Chothia C., Corbani L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,

SQ SEQUENCE 521 AA; 59000 MW; 21D62F370EB33E7D CRC64;
 Query Match 92.7%; Score 255; DB 1; Length 521;
 Best Local Similarity 88.2%; Pred. No. 1.8e-24;
 Matches 45; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 GVNPTAQSSGLSGSIIYALCNCFYPAATGLYVDQVLYHYCVVDPEATA 51
 |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
 Db 194 GVAPTAQASGSMGSIYMLCNCFYPPGTLGYDQVLYHYCVVDPEATA 244
 |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||

RESULT 4
 Q6P6T5 PRELIMINARY; PRT; 523 AA.
 AC Q6P6T5
 DT 03-JUL-2004 (TrEMBLrel. 27, Created)
 DT 03-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 03-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Occludin.
 GN Name:Occln;
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RS SEQUENCE FROM N.A.
 RP
 RC TISSUE=Prostate;
 RC MEDLINE=22398257; PubMed=12477932; DOI=10.1073/pnas.2426038999;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haefl F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield V.S.,
 RA Krzywinski M.I., Skalka U., Smillus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RS SEQUENCE FROM N.A.
 RP
 RC TISSUE=Prostate;
 RC Strausberg R.;
 RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
 RL EMBL; BC062037; AAH62037.1; -.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0005923; C:tight junction; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR008253; Marvel.
 DR InterPro; IPR002958; Occludin.
 DR RefSeq; IPR010844; Occludin_ELL.
 DR Pfam; PF01284; MARVEL; 1.
 DR Pfam; PF07303; Occludin_ELL; 1.
 DR PRINTS; PR01258; OCCLUDIN.
 DR
 SQ SEQUENCE 523 AA; 59186 MW; A31E007B58AEASB5 CRC64;

Query Match 91.3%; Score 251; DB 2; Length 523;
 Best Local Similarity 86.3%; Pred. No. 5.9e-24;
 Matches 44; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```
RESULT 5
Q92303 ID Q92303 PRELIMINARY; PRT; 523 AA.
AC Q92303
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Occludin.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SD; TISSUE=Liver;
RA Kokai Y., Kuwahara K., Atsumi S., Mori M.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016425; BAA36681.1; -.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0005923; C: tight junction; IEA.
DR InterPro; IPR008253; Marvel.
DR InterPro; IPR002358; Occludin.
DR Pfam; PF01284; MARVEL; 1.
DR Pfam; PF07303; Occludin_ELL; 1.
DR PRINTS; PR01258; OCCLUDIN.
SQ SEQUENCE 523 AA; 59214 MW; 5E1E007568AE5B1 CRC64;

Query Match 91.3%; Score 251; DB 2; Length 523;
Best Local Similarity 86.3%; Pred. No. 5.9e-24;
Matches 44; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GVNPTAQSGLSYGSIYALCNQFYTPAATGLYVDQYLHYCVVDPEAIA 51
Db 196 GVNPTAQSGLSYGSIYALCNQFYTPAATGLYVDQYLHYCVVDPEAIA 246

RESULT 6
OCLN CANFA ID OCLN CANFA STANDARD; PRT; 521 AA.
AC Q28269;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Occludin.
GN Name=OCLN;
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=96181088; PubMed=8601611; DOI=10.1083/jcb.133.1.43;
RA Ando-Akatsuka Y., Saitou M., Hirase T., Kishi M., Sakakibara A.,
RA Itoh M., Yonemura S., Furuse M., Tsukita S.;
RT "Interspecies diversity of the occludin sequence: cDNA cloning of
RT human, mouse, dog, and rat-kangaroo homologues.";
RL J. Cell Biol. 133:43-47 (1996).
RN [2]
RP PHOSPHORYLATION.
RX MEDLINE=97327764; PubMed=9182670; DOI=10.1083/jcb.137.6.1393;
RA Sakakibara A., Furuse M., Saitou M., Ando-Akatsuka Y., Tsukita S.;
RT "Possible involvement of phosphorylation of occludin in tight junction
RT formation.";
RL J. Cell Biol. 137:1393-1401 (1997).
CC -! FUNCTION: May play a role in the formation and regulation of the
CC tight junction (TJ) paracellular permeability barrier. Interacts
CC with ZO-1.
CC -! SUBUNIT: Interacts with VAPA (By similarity).
CC -! SUBCELLULAR LOCATION: Integral membrane protein.
```

```
CC -! TISSUE SPECIFICITY: Localized at tight junctions of both
CC epithelial and endothelial cells.
CC -! DOMAIN: The C-terminal is cytoplasmic and is important for
CC interaction with ZO-1. Necessary for the tight junction
CC localization. Involved in the regulation of the permeability
CC barrier function of the tight junction (By similarity).
CC -! PTM: Phosphorylated. Less-phosphorylated forms are found in
CC basolateral membrane, cytosol and tight junction. More-heavily
CC phosphorylated forms are concentrated exclusively in tight
CC junction.
CC -! SIMILARITY: Belongs to the ELL / occludin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U49221; AAC48582.1; -.
CC InterPro; IPR008253; Marvel.
CC InterPro; IPR002958; Occludin.
CC InterPro; IPR010844; Occludin_ELL.
CC Pfam; PF01284; MARVEL; 1.
CC Pfam; PF07303; Occludin_ELL; 1.
CC PRINTS; PR01258; OCCLUDIN.
CC Coiled coil; Phosphorylation; Tight junction; Transmembrane.
FT DOMAIN 1 66
FT TRANSMEM 67 89
FT DOMAIN 90 134
FT TRANSMEM 135 159
FT DOMAIN 160 169
FT TRANSMEM 170 194
FT DOMAIN 195 242
FT TRANSMEM 243 264
FT DOMAIN 265 521
FT DOMAIN 92 130
FT DOMAIN 308 311
FT DOMAIN 424 488
FT SEQUENCE 521 AA; 59275 MW; 2875E59F8F0A1FFA CRC64;

Query Match 88.7%; Score 244; DB 1; Length 521;
Best Local Similarity 86.3%; Pred. No. 4.7e-23;
Matches 44; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GVNPTAQSGLSYGSIYALCNQFYTPAATGLYVDQYLHYCVVDPEAIA 51
Db 195 GVNPTAQSGLSYGSIYALCNQFYTPAATGLYVDQYLHYCVVDPEAIA 245

RESULT 7
Q9NOW3 ID Q9NOW3 PRELIMINARY; PRT; 560 AA.
AC Q9NOW3;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Occludin IB.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=20143583; PubMed=10679019;
RA Muresan Z., Paul D.L., Goodenough D.A.;
RT "Occludin IB, a variant of the tight junction protein occludin.";
RL Mol. Biol. Cell 11:627-634 (2000).
DR EMBL; AF246976; AAF62172.1; -.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0005923; C: tight junction; IEA.
```


DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR008253; Marvel.
 DR InterPro; IPR002958; Occludin.
 DR InterPro; IPR010844; Occludin_ELL.
 DR Pfam; PF01284; MARVEL; 1.
 DR Pfam; PF07303; Occludin_ELL; 1.
 DR PRINTS; PR01258; OCCLUDIN.
 SQ SEQUENCE 560 AA; 63450 MW; D80B178D80AAB12D CRC64;
 Query Match 88.7%; Score 244; DB 2; Length 560;
 Best Local Similarity 86.3%; Pred. No. 5.1e-23;
 Matches 44; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 1 GVNPTAQSGSLYSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOBAIA 51
 DB 234 GVNPTAQSGSLYSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOBAIA 284
 RESULT 8
 OCLN_XENLA STANDARD; PRT; 493 AA.
 AC Q9PUN1;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Occludin.
 GN Name=OCLN;
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A., AND PHOSPHORYLATION SITES THR-375 AND SER-379.
 RC TISSUE=Ovary;
 RX MEDLINE=99421641; PubMed=10491082;
 RA Cordenonsi M., Turco F., D'Atri F., Hammar E., Martinucci G.,
 RA Meggio F., Citi S.;
 RT "Xenopus laevis occludin. Identification of in vitro phosphorylation
 RT sites by protein kinase CK2 and association with cingulin.";
 RL Eur. J. Biochem. 264:374-384(1999).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=98034414; PubMed=9365283;
 RA Cordenonsi M., Mazzon E., De Rigo L., Baraldo S., Meggio F., Citi S.;
 RT "Occludin dephosphorylation in early development of Xenopus laevis.";
 RL J. Cell Sci. 110:3131-3139(1997).
 CC -!- FUNCTION: Probably plays a role in the formation and regulation of
 CC the tight junction (TJ) paracellular permeability barrier.
 CC -!- SUBUNIT: Interacts in vitro with cingulin, possibly directly.
 CC Interacts with ZO-1 (By similarity)
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: Localized at tight junctions of both
 CC epithelial and endothelial cells.
 CC -!- DEVELOPMENTAL STAGE: A maternally synthesized protein. Found in
 CC granules in the peripheral cytoplasm in the fertilized egg, it
 CC localizes first to the basolateral membrane, then to tight
 CC junctions after cingulin and ZO-1. Nascent tight junctions are in
 CC place by the two-cell stage. The maternal form is more highly
 CC phosphorylated than the form detected in later developmental
 CC stages.
 CC -!- DOMAIN: The C-terminus is cytoplasmic and is important for
 CC interaction with ZO-1. Necessary for the tight junction
 CC localization. Involved in the regulation of the permeability
 CC barrier function of the tight junction (By similarity).
 CC -!- PTM: Phosphorylated.
 CC -!- SIMILARITY: Belongs to the ELL / occludin family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL; AF170275; AAD53725.1; -
 DR InAcct; Q9PUN1; -
 DR InterPro; IPR008253; Marvel.
 DR InterPro; IPR002958; Occludin.
 DR InterPro; IPR010844; Occludin_ELL.
 DR Pfam; PF01284; MARVEL; 1.
 DR Pfam; PF07303; Occludin_ELL; 1.
 DR PRINTS; PR01258; OCCLUDIN.
 KW Coiled coil; Phosphorylation; Tight junction; Transmembrane.
 FT DOMAIN 1 47
 FT TRANSMEM 48 70
 FT POTENTIAL
 FT DOMAIN 71 116
 FT EXTRACELLULAR (Potential).
 FT TRANSMEM 117 141
 FT POTENTIAL
 FT DOMAIN 142 151
 FT CYTOPLASMIC (Potential).
 FT TRANSMEM 152 176
 FT POTENTIAL
 FT DOMAIN 177 224
 FT EXTRACELLULAR (Potential).
 FT TRANSMEM 225 246
 FT POTENTIAL
 FT DOMAIN 247 493
 FT CYTOPLASMIC (Potential).
 FT DOMAIN 396 428
 FT COILED COIL (Potential).
 FT MOD_RES 375 375
 FT PHOSPHOTHREONINE (by CK2) (in vitro).
 FT MOD_RES 379 379
 FT PHOSPHOSERINE (by CK2) (in vitro).
 SQ SEQUENCE 493 AA; 55111 MW; 9694CD302BEBDE CRC64;
 Query Match 74.2%; Score 204; DB 1; Length 493;
 Best Local Similarity 68.6%; Pred. No. 7.2e-18;
 Matches 35; Conservative 10; Mismatches 6; Indels 0; Gaps 0;
 QY 1 GVNPTAQSGSLYSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOBAIA 51
 DB 177 GVNPTAQSGSAFYTVQISVCNQFYSPVQTVGVNQYLYHYCVVDPOBAIA 227
 RESULT 9
 Q66J48
 ID Q66J48 PRELIMINARY; PRT; 519 AA.
 AC Q66J48;
 DT 25-OCT-2004 (TRENBLrel. 28, Created)
 DT 25-OCT-2004 (TRENBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TRENBLrel. 28, Last annotation update)
 DE MGC81923 protein.
 GN Name=MGC81923;
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
 RA Richardson P.;
 RA "Genetic and genomic tools for Xenopus research: The NIH Xenopus
 RA initiative.";
 RL Dev. Dyn. 225:384-391(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S.J., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

DR	Pfam; PF07303; Occludin_ELL; 1.
DR	PRINTS; PR01258; OCCLUDIN.
KW	Cooled coil; Phosphorylation; Tight junction; Transmembrane.
FT	DOMAIN 1 51
FT	FT TRANSMEM 52 74
FT	FT DOMAIN 51 74
FT	FT TRANSMEM 75 112
FT	FT DOMAIN 113 137
FT	FT TRANSMEM 138 147
FT	FT TRANSMEM 148 172
FT	FT DOMAIN 173 222
FT	FT TRANSMEM 223 244
FT	FT TRANSMEM 245 489
FT	FT DOMAIN 76 107
FT	FT DOMAIN 407 434
SQ	SEQUENCE 489 AA; 54075 MW; 4F0CA45A41094860 CRC64;
 Query Match	
Best Local Similarity 59.3%; Score 163; DB 1; Length 489;	
Matches 33; Conservative 8; Mismatches 9; Indels 4; Gaps	
QY	1 GVNPTA---QQSGSLYGSIYALCNQCYTPPAATGLVYDVQYLHYCVVDPQEATA 51 : : : : : : :
DB	173 GWNPRAGLGSSGSLYINQLMCLNQNMSPVAGGI-MNQLYHYCNDVPQEAVA 225 : : : : : : :
 RESULT 13	
Q6NX99	PRELIMINARY; PRT; 500 AA.
ID Q6NX99	
AC Q6NX99	
DT 05-JUL-2004 (TrEMBLrel. 27, Created)	
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)	
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)	
DE Hypothetical protein (Fragment).	
OS Brachydanio rerio (Zebrafish) (Danio rerio).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes	
OC Cyprinidae; Danio.	
ON NCBI_TaxID=7955;	
RX [1]	
RP SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;	
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,	
RA Klausner R.D., Collins F.S., Wagner L., Shennen C.M., Schuler G.D.,	
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,	
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,	
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,	
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.	
RA Brownstein M.J., Usdin T.B., Tohiyuki S., Carninci P., Prange C.,	
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.	
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,	
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.	
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,	
RA Fahay J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.	
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,	
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,	
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.	
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,	
RA Jones S.J., Warra M.A.;	
RA "Generation and initial analysis of more than 15,000 full-length hum	
RT and mouse cDNA sequences.";	
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).	
RP SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.	
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.	
RL EMBL; BC067178; AAA67178.1; .	
GO GO; GO:0016021; C:integral to membrane; IEA.	
DR GO; GO:0005923; C:tight junction; IEA.	
DR GO; GO:0005198; P:structural molecule activity; IEA.	
DR InterPro; IPR008253; Marvel.	
DR InterPro; IPR002958; Occludin.	

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2005, 09:16:13 ; Search time 64.67 Seconds
(without alignments)
565.610 Million cell updates/sec

Title: US-09-891-064a-2_COPY_33_522
Perfect score: 2590
Sequence: 1 MHVRPMLSQPAYFYPEDEI.....LKSLSHIKMWGVDRQKT 490

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/aaa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/aaa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/aaa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/aaa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/aaa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/aaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2590	100.0	522	3	US-09-142-732-2
2	2590	100.0	522	4	US-08-945-826-2
3	2590	100.0	522	4	US-09-197-503-2
4	2410.5	93.1	521	4	US-08-945-826-4
5	2410.5	93.1	521	4	US-09-197-503-4
6	2335.5	90.2	521	4	US-08-945-826-6
7	2335.5	90.2	521	4	US-09-197-503-6
8	566	21.9	108	3	US-09-026-343-8
9	566	21.9	108	3	US-09-362-871-8
10	263	10.2	48	3	US-09-222-373-5
11	263	10.2	48	3	US-08-996-679-59
12	263	10.2	48	3	US-09-001-511-5
13	263	10.2	48	3	US-09-113-977C-69
14	263	10.2	48	3	US-09-510-616-5
15	263	10.2	48	3	US-09-248-074-78
16	263	10.2	48	3	US-09-187-859-62
17	263	10.2	48	4	US-09-458-870-78
18	263	10.2	48	4	US-09-351-048A-69
19	263	10.2	48	4	US-09-839-542B-62
20	263	10.2	48	4	US-09-450-073-5
21	263	10.2	48	4	US-10-193-653-69
22	261	9.7	47	3	US-08-939-853A-15
23	243	9.4	48	3	US-09-222-373-6
24	243	9.4	48	3	US-09-001-511-6
25	243	9.4	48	3	US-09-510-616-6
26	243	9.4	48	4	US-09-450-073-6
27	232	9.0	48	3	US-09-222-373-7

28	232	9.0	48	3	US-09-001-511-7	Sequence 7, Appli
29	232	9.0	48	3	US-09-510-616-7	Sequence 7, Appli
30	232	9.0	48	4	US-09-450-073-7	Sequence 7, Appli
31	229.5	8.9	560	4	US-09-949-016-8174	Sequence 8174, Ap
32	229.5	8.9	621	3	US-09-026-343-7	Sequence 7, Appli
33	229.5	8.9	621	3	US-09-362-871-7	Sequence 7, Appli
34	229.5	8.9	621	4	US-09-943-016-6737	Sequence 6737, Ap
35	221.5	8.6	51	3	US-09-222-373-46	Sequence 46, Appl
36	221.5	8.6	51	3	US-09-001-511-46	Sequence 46, Appl
37	221.5	8.6	51	3	US-09-510-616-46	Sequence 46, Appl
38	221.5	8.6	51	4	US-09-450-073-46	Sequence 46, Appl
39	219	8.5	640	3	US-09-026-343-2	Sequence 2, Appli
40	219	8.5	640	3	US-09-362-871-2	Sequence 2, Appli
41	152	5.9	50	3	US-09-222-373-8	Sequence 8, Appli
42	152	5.9	50	3	US-09-001-511-8	Sequence 8, Appli
43	152	5.9	50	3	US-09-510-616-8	Sequence 8, Appli
44	152	5.9	50	4	US-09-450-073-8	Sequence 8, Appli
45	145	5.6	464	4	US-09-270-767-61287	Sequence 61287, A

ALIGNMENTS

RESULT 1
US-09-142-732-2
; Sequence 2, Application US/09142732
; Patent No. 6252045
; GENERAL INFORMATION:
; APPLICANT: James M. Anderson
; APPLICANT: Christina M. Van Itallie
; TITLE OF INVENTION: Human Occludin, Its Uses and
; TITLE OF INVENTION: Enhancement of Drug Absorption Using Occludin Inhibitors
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yale University Medical School
; ADDRESSEE: Section of Digestive Diseases
; ADDRESSEE: Department of Internal Medicine
; STREET: 333 Cedar Street, LCI 105
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: United States of America
; ZIP: 06520-8057
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" 1.44 Mb diskette
; COMPUTER: IBM PC
; OPERATING SYSTEM: MS DOS
; SOFTWARE: Word Processing
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/142.732
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/05809
; FILING DATE: March 14, 1997
; APPLICATION NUMBER: U.S. 60/013,625
; FILING DATE: March 15, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary M. Krinsky
; REGISTRATION NUMBER: 32423
; REFERENCE/DOCKET NUMBER: 1751-P0016B.PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 203-324-6155
; TELEFAX: 203-327-1096
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 522
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: polypeptide
; FRAGMENT TYPE: complete sequence
; FEATURE:

RESULT 4
US-08-945-826--4
; Sequence 4, Application US/08945826
; Patent No. 6489460
; GENERAL INFORMATION:
; APPLICANT: TSUKITA, Shoichiro
; TITLE OF INVENTION: HUMAN ADHESION MOLECULE OCCLUDIN
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church

RESULT 5
US-09-197-503-4
; Sequence 4, Application US/09137503
; Patent No. 6559286
; GENERAL INFORMATION:

Db 390 KRTDPHYETDVTGGESCELEEDWVREYPPITSDQQQLYKRNFDAGLQYKSLQAE 449
QY 419 DEINKELSLDKELDDYRESEYMAAADEYNRLKQVKSADYKSKQNHCKOLKSLSHI 478
Db 450 DDVNKELSLDKELDDYRESEYMAAADEYNRLKQVKSADYKSKRNYCKOLKSLSHI 509
QY 479 KRMVGDYDRQK 489
Db 510 KRMVGDYDRRK 520

RESULT 7

US-09-197-503-6
; Sequence 6, Application US/09197503
; Patent No. 6559286
; GENERAL INFORMATION:
; APPLICANT: TSUKITA, Shoichiro
; TITLE OF INVENTION: HUMAN ADHESION MOLECULE OCCLUDIN
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/197,503
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/945,826
; FILING DATE: 05-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: PARACT, C. Joseph
; REGISTRATION NUMBER: 32,350
; REFERENCE/DOCKET NUMBER: 0425-0660P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 521 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-197-503-6

Query Match 90.2%; Score 2335.5; DB 4; Length 521;
Best Local Similarity 90.0%; Pred. No. 4.4e-199;
Matches 442; Conservative 20; Mismatches 24; Indels 5; Gaps 4;
QY 1 MHVRPMLQPAYSPFPEDEILHFYKWTSPPGVIRILSMLIIYVNCIAIFACVASTLAWDRG 60
Db 33 MHVRPMLQPAYSPFPEDEILHFYKWTSPPGVIRILSMLIIYVNCIAIFACVASTLAWDRG 92
QY 61 YGTSLLGSGVGPYGGSGFGSGYGVGY -GYGYGGYTDPPRAAKGFLMAAFCFIA 119
Db 93 YGTGLFGSLNYPY --SGFG-YGGYGGYGGYGVGYGTYDPPRAAKGFLMAAFCFIA 149
QY 120 ALVIFVTSVIRSEMRTRYLLSVIIVSAILGIMVFIATIVYIMGVNPTAQSSGSLYGSQ 179
Db 150 SLVIFVTSVIRSGMRTRYLLSVIIVSAILGIMVFIATIVYIMGVNPTAQSSGSMYGSQ 209
QY 180 IYALCNQFVTPAATGLYDQYLHYHVCVDPQBAIAIVLGFMIIVAFALIIPFAVTRRM 239
Db 210 IYMICNQFVTPGGLYDQYLHYHVCVDPQBAIAIVLGFMIIVAFALIIPFAVTRRM 269

QY 240 DRYDKSNILWKEHIYDQPPNVEEWWKVNVSAGTQVFPSPDYVERVDSPPMAYSSNGKV 299
Db 270 DRYDKSNILWKEHIYDQPPNVEEWWKVNVSAGTQVFPSPDYAERVDSPPMAYSSNGKV 329
QY 300 NDKRFPYESSYKSTP -VPEVVQELPLTSPVDDFRQPRYSSGGNFETTPSKRAPAKGRAGRS 358
Db 330 NGRSYSPESFYKSTPLVPEVAQEIPLTLSVDDFRQPRYSSNGNLETTPSKRAFTKGKAGK 389
QY 359 KRTEODHYETDVTGGESCELEEDWVREYPPITSDQQQLYKRNFDAGLQYKSLQAE 418
Db 390 KRTDPHYETDVTGGESCELEEDWVREYPPITSDQQQLYKRNFDAGLQYKSLQAE 449
QY 419 DEINKELSLDKELDDYRESEYMAAADEYNRLKQVKSADYKSKQNHCKOLKSLSHI 478
Db 450 DDVNKELSLDKELDDYRESEYMAAADEYNRLKQVKSADYKSKRNYCKOLKSLSHI 509
QY 479 KRMVGDYDRQK 489
Db 510 KRMVGDYDRRK 520

RESULT 8

US-09-026-343-8
; Sequence 8, Application US/09026343
; Patent No. 6008018
; GENERAL INFORMATION:
; APPLICANT: DUAN, D. ROXANNE
; APPLICANT: SHILATIFARD, ALI
; APPLICANT: CONAWAY, JOAN W.
; APPLICANT: CONAWAY, RONALD C.
; TITLE OF INVENTION: ELL2, A New Member of an ELL Family of
; TITLE OF INVENTION: RNA Polymerase II Elongation Factors
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/026,343
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/038,447
; FILING DATE: 19-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: GOLDSTEIN, JORGE A.
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 1488.0880001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 372-2540
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
US-09-026-343-8

Query Match 21.9%; Score 566; DB 3; Length 108;
Best Local Similarity 100.0%; Pred. No. 1e-42;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 383 DWIREYPPITSDQOQLYKRNFDTGLQYKSLQSELDEINKLSRLDKELDDYRESEY 442
|||||
Db 1 DWIREYPPITSDQOQLYKRNFDTGLQYKSLQSELDEINKLSRLDKELDDYRESEY 60
|||||
QY 443 MAADEYNRLKQVGSADYKSKKNHCKQLKSLSHIKKMGVDYDRQKT 490
|||||
Db 61 MAADEYNRLKQVGSADYKSKKNHCKQLKSLSHIKKMGVDYDRQKT 108
|||||

RESULT 9

US-09-362-871-8
; Sequence 8, Application US/09362871
; Patent No. 6379923
; GENERAL INFORMATION:
; APPLICANT: DUAN, D. ROXANNE
; APPLICANT: SHILATIFARD, ALI
; APPLICANT: CONAWAY, JOAN W.
; APPLICANT: CONAWAY, RONALD C.
; TITLE OF INVENTION: ELL2, A New Member of an ELL Family of
; TITLE OF INVENTION: RNA Polymerase II Elongation Factors
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/362,871
; FILING DATE:
; CLASSIFICATION:

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/026,343
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: GOLDSTEIN, JORGE A.
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 1488.0880001

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 372-2540
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant

; MOLECULE TYPE: protein

US-09-362-871-8

Query Match 21.9%; Score 566; DB 3; Length 108;
Best Local Similarity 100.0%; Pred. No. 1e-42;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 383 DWIREYPPITSDQOQLYKRNFDTGLQYKSLQSELDEINKLSRLDKELDDYRESEY 442
|||||
Db 1 DWIREYPPITSDQOQLYKRNFDTGLQYKSLQSELDEINKLSRLDKELDDYRESEY 60
|||||
QY 443 MAADEYNRLKQVGSADYKSKKNHCKQLKSLSHIKKMGVDYDRQKT 490
|||||
Db 61 MAADEYNRLKQVGSADYKSKKNHCKQLKSLSHIKKMGVDYDRQKT 108
|||||

RESULT 10

US-09-222-373-5
; Sequence 5, Application US/09222373
; Patent No. 6110747

; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING TISSUE
; TITLE OF INVENTION: PERMEABILITY
; FILE REFERENCE: 100086.405C1
; CURRENT APPLICATION NUMBER: US/09/222,373
; CURRENT FILING DATE: 1998-12-29
; EARLIER APPLICATION NUMBER: US 09/001,511
; EARLIER FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-222-373-5

Query Match 10.2%; Score 263; DB 3; Length 48;
Best Local Similarity 100.0%; Pred. No. 2.7e-16;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQLYHYCVVDPOE 211
|||||
Db 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQLYHYCVVDPOE 48
|||||

RESULT 11

US-08-996-679-59
; Sequence 59, Application US/08996679
; Patent No. 6169071
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING
; TITLE OF INVENTION: CELL ADHESION
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/996,679
; APPLICATION NUMBER: US/08/996,679
; FILING DATE: 23-DEC-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 100086.401C1

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

Query Match 10.2%; Score 263; DB 3; Length 48;
Best Local Similarity 100.0%; Pred. No. 2.7e-16;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 GVNPTAQSSGLSGSIQIYALCNQFYTPAATGLYVDQYLYHYCVDPQE 211
Db 1 GVNPTAQSSGLSGSIQIYALCNQFYTPAATGLYVDQYLYHYCVDPQE 48

RESULT 12

US-09-001-511-5
; Sequence 5, Application US/09001511
; Patent No. 624864
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING
; TITLE OF INVENTION: TISSUE PERMEABILITY
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/001,511
; FILING DATE: 31-DEC-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 100086.405
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-001-511-5

Query Match 10.2%; Score 263; DB 3; Length 48;
Best Local Similarity 100.0%; Pred. No. 2.7e-16;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 GVNPTAQSSGLSGSIQIYALCNQFYTPAATGLYVDQYLYHYCVDPQE 211
Db 1 GVNPTAQSSGLSGSIQIYALCNQFYTPAATGLYVDQYLYHYCVDPQE 48

RESULT 13

US-09-113-977C-69
; Sequence 69, Application US/09113977C
; Patent No. 627824
; GENERAL INFORMATION:
; APPLICANT: Doherty, Patrick
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING ADHESION MOLECULE
; TITLE OF INVENTION: FUNCTION
; FILE REFERENCE: 100086.403
; CURRENT APPLICATION NUMBER: US/09/113,977C
; CURRENT FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 69
; LENGTH: 48
; TYPE: PRT

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Solid Phase
; OTHER INFORMATION: Synthesis
US-09-113-977C-69

Query Match 10.2%; Score 263; DB 3; Length 48;
Best Local Similarity 100.0%; Pred. No. 2.7e-16;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 GVNPTAQSSGLSGSIQIYALCNQFYTPAATGLYVDQYLYHYCVDPQE 211
Db 1 GVNPTAQSSGLSGSIQIYALCNQFYTPAATGLYVDQYLYHYCVDPQE 48

RESULT 14

US-09-510-616-5
; Sequence 5, Application US/09510616
; Patent No. 6310177
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Symonds, James Matthew
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING
; TITLE OF INVENTION: TISSUE PERMEABILITY
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED IP LAW GROUP PLLC
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/510,616
; FILING DATE: 22-Feb-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Christiansen, William T.
; REGISTRATION NUMBER: 44,614
; REFERENCE/DOCKET NUMBER: 100086.405C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-510-616-5

Query Match 10.2%; Score 263; DB 3; Length 48;
Best Local Similarity 100.0%; Pred. No. 2.7e-16;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 GVNPTAQSSGLSGSIQIYALCNQFYTPAATGLYVDQYLYHYCVDPQE 211
Db 1 GVNPTAQSSGLSGSIQIYALCNQFYTPAATGLYVDQYLYHYCVDPQE 48

RESULT 15

US-09-248-074-78
; Sequence 78, Application US/09248074
; Patent No. 6346512
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.

```

; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING CELL ADHESION
; FILE REFERENCE: 100086.401C5
; CURRENT APPLICATION NUMBER: US/09/248,074
; CURRENT FILING DATE: 1999-02-10
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 78
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Occludin cell
; OTHER INFORMATION: adhesion recognition sequence and flanking amino
; OTHER INFORMATION: acids
US-09-248-074-78

```

```

Query Match      10.2%; Score 263; DB 3; Length 48;
Best Local Similarity 100.0%; Pred. No. 2.7e-16;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      164 GVNPTAQSSGLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVDPQE 211
          |||||
Db       1 GVNPTAQSSGLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVDPQE 48

```

Search completed: April 1, 2005, 10:04:44
Job time : 65.67 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2005, 09:16:14 ; Search time 26.6497 Seconds
(without alignments)
725.636 Million cell updates/sec

Title: US-09-891-064A-2_COPY_89_138

Perfect score: 293
Sequence: 1 WDRGYGTSLLGSGVPGYGG.....GYGYGYGYGYGYDTPRAAK 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	293	100.0	508	8	Abm84985 Human dia
2	293	100.0	508	8	Abm84986 Human dia
3	293	100.0	522	2	Aaw36052 Human occ
4	293	100.0	522	3	Aaw34638 Human occ
5	293	100.0	522	3	Aab35731 Human occ
6	293	100.0	522	6	Abj37076 Human bre
7	293	100.0	522	7	Add46545 Human pro
8	293	100.0	522	8	Adi47189 Human pro
9	262.5	89.6	505	2	Aaw36053 Canine oc
10	233	79.5	523	7	Add46543 Rat Prote
11	214	73.0	521	2	Aaw36054 Mouse occ
12	141	48.1	24	2	Aaw34640 Human occ
13	128	43.7	24	2	Aaw34639 Human occ
14	120.5	41.1	62	7	Adj92285 Mouse hai
15	120.5	41.1	141	7	Adj92275 Mouse hai
16	118	40.3	90	7	Adj92185 Human hai
17	118	40.3	139	4	Abg27994 Novel hum
18	116.5	39.8	81	4	Abp69638 Human pol
19	116.5	39.8	81	5	Abp69638 Human pol
20	116.5	39.8	81	7	Adj92187 Human hai
21	115.5	39.4	209	4	Abb61235 Drosophil
22	112.5	38.4	78	7	Adj92267 Mouse hai
23	112.5	38.4	78	7	Adj92265 Mouse hai
24	111	37.9	440	4	Abb60862 Drosophil
25	110.5	37.7	63	7	Adj92217 Human hai

26	110.5	37.7	87	7	ADJ92277 Mouse hai
27	110.5	37.7	87	7	ADJ92273 Mouse hai
28	110.5	37.7	84	7	ADJ92283 Mouse hai
29	108.5	37.0	134	6	ABU09590 Tick infe
30	107.5	36.7	119	7	ADD26165 Protein r
31	107.5	36.7	152	4	ABE69237 Drosophil
32	106.5	36.3	128	7	ADJ92269 Mouse hai
33	106.5	36.3	154	2	AAJ13500 Tissue ce
34	106.5	36.3	154	5	AAW50381 Tick infe
35	106.5	36.3	154	6	ABU09592 Tick infe
36	106	36.2	83	7	ADJ92211 Human hai
37	105.5	36.0	79	7	ADJ92213 Human hai
38	105	35.8	80	7	ADJ92281 Mouse hai
39	105	35.8	212	4	AAJ75147 Human col
40	105	35.8	218	3	AAW57093 Human pro
41	105	35.8	341	2	AAW26553 Human het
42	105	35.8	341	2	AAW26553 Human het
43	105	35.8	341	8	ABM80856 Tumour-as
44	105	35.8	353	2	AAW26552 Human het
45	105	35.8	353	2	AAW50921 Amino aci

ALIGNMENTS

RESULT 1
ABM84985
ID ABM84985 standard; protein; 508 AA.
XX
AC ABM84985;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human diagnostic and therapeutic pprotein SEQ ID NO:5234.
XX
KW gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX
OS Homo sapiens.
XX
PN WO2004023973-A2.
XX
PD 25-MAR-2004.
XX
PF 12-SEP-2003; 2003WO-US028227.
XX
PR 12-SEP-2002; 2002US-0410259P.
PR 12-SEP-2002; 2002US-0410260P.
XX
(INCY-) INCYTE CORP.

Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F; Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV; Mooney EM, Delegeane AM, Panesar IS, Barville SC, Reddy TP; Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstein EH; Peralta CH, Anderson SB, Rioux P, Shen EU, Wu MC, Stuve LL; Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES; Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D; Patury S, Shi X, Suarez CJ;
WPI; 2004-329368/30.
N-PSDB; ACN43637.
New diagnostic and therapeutic polynucleotides and polypeptides, useful in diagnosing a condition, disease or disorder associated with human molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or in gene mapping.
Claim 27; Page; 190pp; English.
The invention relates to novel diagnostic and therapeutic polynucleotides selected from one of the 2722 sequences defined in the specification. A polynucleotide of the invention may have a use in gene therapy. The human diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be

Qy 1 WDRGYGTSLLGGSVGPYPYGGSGFGSYGSYGSGYGYGYGGYTDPRAAK 50
|||
Dp 89 WDRGYGTSLLGGSVGPYPYGGSGFGSYGSYGSGYGYGYGGYTDPRAAK 138

KW occludin; claudin; sexual dysfunction; male erectile sexual dysfunction;
XX female sexual dysfunction; human.
XX Homo sapiens.
XX WO2004003145-A2.
XX 08-JAN-2004.
XX 24-JUN-2003; 2003WO-US019994.
XX 28-JUN-2002; 2002US-0392512P.
XX (NAST-) NASTECH PHARM CO INC.
XX Quay SC;
XX WPI; 2004-091087/09.
XX Composition comprising biologically active agent and permeabilizing
PT peptide that reversibly enhances mucosal epithelial paracellular
PT transport by modulating epithelial junctional structure and/or physiology
PT in mammal.
XX Disclosure; SEQ ID NO 900; 426pp; English.
XX This invention relates to a novel composition comprising a biologically
CC active agent and mucosal delivery-enhancing effective amount of
CC permeabilizing peptide that reversibly enhances mucosal epithelial
CC paracellular transport by modulating epithelial junctional structure
CC and/or physiology in a mammal. The agent of the invention inhibits
CC homotypic binding of epithelial membrane adhesive protein chosen
CC functional adhesion molecule (JAM), occludin and claudin. The
CC biologically active agent is effective for treatment of sexual
CC dysfunction, for example male erectile sexual dysfunction or female
CC sexual dysfunction. The present sequence is that of the human occludin
CC protein which is related to the invention.
XX Sequence 522 AA;
Query Match 100.0%; Score 293; DB 8; Length 522;
Best Local Similarity 100.0%; Pred. No. 1.1e-22;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 WDRGVTSLGGSGVGPYGGSGFGSGYGVGYGVGYGTYDPPRAAK 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 89 WDRGVTSLGGSGVGPYGGSGFGSGYGVGYGVGYGTYDPPRAAK 138
RESULT 9
AAW36053
ID AAW36053 standard; protein; 505 AA.
XX AAW36053;
XX 25-MAR-1998 (first entry)
XX Canine occludin protein sequence.
XX Occludin; adhesion molecule; membrane tight junction;
KW occludin localisation; membrane; occludin expression;
KW blood-brain barrier disorder.
XX Canis sp.
XX WO9732982-A1.
XX 12-SEP-1997.
XX 05-MAR-1997; 97WO-JP000665.
XX 07-MAR-1996; 96JP-00049880.
XX 12-DEC-1996; 96JP-00331944.

XX (EISA) EISAI CO LTD.
XX Tsukita S;
XX WPI; 1997-470546/43.
XX N-PSDB; AAT97973.
XX DNA encoding human, dog and mouse occludin(s) - useful for screening for
PT substances influencing occluding expression.
XX Claim 4; Page 17-20; 36pp; Japanese.
XX The present sequence represents canine occludin, a novel adhesion
CC molecule which is a constituent protein of membrane tight junctions. The
CC present sequence is encoded by AAT97973, which encodes an extra 16 amino
CC acid residues in the correct reading frame, which are not shown in the
CC specification. Antibodies raised against the protein are used to assay
CC for occludin in samples, and for the study of occludin localisation in
CC membranes (e.g. by immunofluorescent cell-staining). The transformants
CC and antibodies are used for screening of substances which potentially
CC influence occludin expression. Therapeutic polynucleotides derived from
CC the DNA encoding the present sequence are used for treatment of disorders
CC involving the blood-brain barrier. Primers that hybridise to the occludin
CC DNA sequence can be used to detect occludin DNA by PCR
XX Sequence 505 AA;
Query Match 89.6%; Score 262.5; DB 2; Length 505;
Best Local Similarity 88.0%; Pred. No. 1.6e-19;
Matches 44; Conservative 4; Mismatches 1; Indels 1; Gaps 1;
QY 1 WDRGVTSLGGSGVGPYGGSGFGSGYGVGYGVGYGTYDPPRAAK 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 89 WDRGVTSLGGSGVGPYGGSGFGSGYGVGYGVGYGTYDPPRAAK 137
RESULT 10
ADD46543
ID ADD46543 standard; protein; 523 AA.
XX ADD46543;
XX 02-DEC-2004 (revised)
DT 29-JAN-2004 (first entry)
XX Rat Protein BAA36681, SEQ ID NO 12224.
XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX Rattus norvegicus.
XX Unidentified.
XX WO2003016475-A2.
XX 27-FEB-2003.
XX 14-AUG-2002; 2002WO-US025765.
XX 14-AUG-2001; 2001US-0312147P.
XX 01-NOV-2001; 2001US-0346382P.
XX 26-NOV-2001; 2001US-0333347P.
XX (GEHO) GEN HOSPITAL CORP.
XX (FARB) BAYER AG.
XX Woolf C, D'urso D, Befort K, Costigan M;
XX WPI; 2003-268312/26.
XX GENBANK; BAA36681.
XX New composition comprising two or more isolated polypeptides, useful for

CC therapy vectors) by increasing absorption across endothelial or
 CC epithelial barriers, i.e. transmucosal or transvascular drug delivery.
 CC Inhibitors include occludin surface loop peptides that inhibit adhesion
 CC and/or barrier properties, or antibodies that interact with occludin or
 CC occludin receptors
 XX
 SQ Sequence 24 AA;

Query Match 48.1%; Score 141; DB 2; Length 24;
 Best Local Similarity 100.0%; Pred. No. 5.5e-08;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 SYGSGYGYGYGYGYGYGYTDP 47
 DB 2 SYGSGYGYGYGYGYGYGYTDP 24
 |||||

RESULT 13
 AAW34639
 ID AAW34639 standard; peptide; 24 AA.
 XX
 AC AAW34639;
 XX
 DT 02-MAR-1998 (first entry)
 XX
 DE Human occludin peptide 1 (aa90-112).
 XX
 KW Occludin; inhibitor; human; drug absorption; drug delivery; gene therapy;
 KW tissue permeability; cell adhesion; antibody.
 XX Homo sapiens.
 OS Synthetic.
 XX
 PN WO9733605-A1.
 XX
 PD 18-SEP-1997.
 XX
 PF 14-MAR-1997; 97WO-US005809.
 XX
 PR 15-MAR-1996; 96US-0013625P.
 XX
 PA (UYFA) UNIV YALE.
 XX
 PI Anderson JM, Van Itallie CM;
 XX
 DR WPI; 1997-470640/43.
 XX
 PT Isolated human occludin protein - useful for increasing drug delivery
 PT across endothelial or epithelial barriers, or for reducing tissue
 PT permeability.
 XX
 PS Example 2; Page 17; 49pp; English.

XX This peptide (#1) comprises amino acids 90-112 of human occludin (see
 CC AAW34638) plus an N-terminal Cys residue to allow conjugation for
 CC antibody production. Peptide #2 (see AAW34640) comprises amino acids 113-
 CC 135 plus N-terminal Cys. Together these contiguous peptides compose the
 CC putative first extracellular loop of the occludin protein. They were used
 CC to raise rabbit polyclonal anti-peptide antibodies. Both peptides are
 CC capable of inhibiting adhesion in occludin-transfected fibroblasts.
 CC Occludin inhibitors can be used to enhance delivery of drugs (or gene
 CC therapy vectors) by increasing absorption across endothelial or
 CC epithelial barriers, i.e. transmucosal or transvascular drug delivery.
 CC Inhibitors include occludin surface loop peptides that inhibit adhesion
 CC and/or barrier properties, or antibodies that interact with occludin or
 CC occludin receptors
 XX
 SQ Sequence 24 AA;
 Query Match 43.7%; Score 128; DB 2; Length 24;
 Best Local Similarity 100.0%; Pred. No. 1.3e-06;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 DRGYTSLGGSVGYGYGSGFG 24
 DB 2 DRGYTSLGGSVGYGYGSGFG 24
 |||||

RESULT 14
 ADJ92285
 ID ADJ92285 standard; protein; 62 AA.
 XX
 AC ADJ92285;
 XX
 DT 06-MAY-2004 (first entry)
 XX
 DE Mouse hair keratin-associated-protein SEQ ID NO:144.
 XX
 KW hair; keratin-associated protein; KAP; mouse; keratin; toiletry;
 KW therapeutic; hair growth promoter; hair disorder.
 XX Mus musculus.
 OS
 PN WO2003042387-A1.
 XX
 PD 22-MAY-2003.
 XX
 PF 13-NOV-2002; 2002WO-JP011851.
 XX
 PR 13-NOV-2001; 2001JP-00348050.
 XX
 PA (UYKE-) UNIV KEIO.
 XX
 PI (NIPR-) JAPAN SOC PROMOTION SCI.
 XX
 PI Kudo J, Shibuya K, Shimizu N;
 XX
 DR WPI; 2003-493307/46.
 DR P-FSDB; ADJ92284.
 XX
 PT DNAs encoding 39 Keratin-associated proteins localized on human
 PT chromosome 21, useful for screening binding and expression modifiers and
 PT as cosmetic and therapeutic agents for hair disorders.
 XX
 PS Example 7; SEQ ID NO 144; 352pp; Japanese.

XX The present invention describes DNAs encoding hair keratin-associated
 CC proteins (KAP) of human origin, which bind to hair keratin. Also
 CC described: (1) DNA encoding KAPs of mouse origin; (2) proteins encoded by
 CC the human and mouse DNA, and their partial peptides, and proteins derived
 CC from them by addition, deletion and/or substitution of one or more amino
 CC acid residues; (3) DNA hybridising to the DNA encoding KAP; (4) peptides
 CC L or F; (5) fusion proteins and peptides containing these proteins and
 CC peptides together with a marker protein or peptide; (6) antibodies to the
 CC proteins and peptides; (7) recombinant proteins and peptides binding to
 CC these antibodies; (8) expression vectors containing the DNA encoding KAP;
 CC (9) host cells transformed by these vectors; (10) non-human animals which
 CC are knockout animals for KAP or which overexpress KAP; (11) screening
 CC substances promoting or inhibiting the binding of KAP to hair keratin, or
 CC promoting or inhibiting the expression of KAP, comprising using KAP or
 CC their partial peptides or cells expressing them; (12) compounds
 CC identified by the screening method; and (13) toiletry and therapeutic
 CC compositions containing these compounds, or containing KAP or their
 CC partial peptides or (I) or (II), expression vectors for them, or host
 CC cells transformed by these vectors. KAP sequences can be used as hair
 CC growth promoters. The KAPs are useful as active ingredients in toiletry
 CC compositions (such as hair and beard growth improvers, hair colourants
 CC and hair conditioners) and in therapeutic compositions for hair
 CC disorders. The present sequence is used in the exemplification of the
 CC present invention.
 XX
 SQ Sequence 62 AA;

Query Match 41.1%; Score 120.5; DB 7; Length 62;
 Best Local Similarity 57.4%; Pred. No. 2e-05;
 Matches 27; Conservative 4; Mismatches 9; Indels 7; Gaps 4;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2005, 09:16:13 ; Search time 137.553 Seconds
(without alignments)
122.943 Million cell updates/sec

Title: US-09-891-064A-2_COPY_196_246
Perfect score: 275
Sequence: 1 GWNPTAQSSGLSGSIYAL.....LYVDQYLHYCVVDPOQAIA 51

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1413372 seqs, 331592847 residues

Total number of hits satisfying chosen parameters: 1413372

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications AA:*
- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
 - 2: /cgn2_6/ptodata/2/pubpaa/CT_NEW_PUB.pep:*
 - 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
 - 6: /cgn2_6/ptodata/2/pubpaa/CTUS_PUBCOMB.pep:*
 - 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
 - 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
 - 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
 - 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
 - 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
 - 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
 - 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
 - 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
 - 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
 - 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
 - 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	275	100.0	522	9	US-09-891-064A-2
2	275	100.0	522	14	US-10-176-847-104
3	275	100.0	522	15	US-10-601-953-900
4	263	95.6	48	9	US-09-769-145-78
5	263	95.6	48	14	US-10-105-008-78
6	263	95.6	48	14	US-10-006-869-62
7	263	95.6	48	14	US-10-193-032-62
8	263	95.6	48	15	US-10-395-032-62
9	263	95.6	48	15	US-10-425-557-78
10	263	95.6	48	15	US-10-412-701-78
11	263	95.6	48	16	US-10-632-678-78
12	263	95.6	48	16	US-10-714-564A-1352
13	263	95.6	48	17	US-10-885-482-5

14	251	91.3	47	10	US-09-778-026-15	Sequence 15, Appl
15	251	91.3	47	16	US-10-648-854-15	Sequence 15, Appl
16	243	88.4	48	17	US-10-885-482-6	Sequence 6, Appl
17	232	84.4	48	17	US-10-885-482-7	Sequence 7, Appl
18	221.5	80.5	51	17	US-10-885-482-46	Sequence 46, Appl
19	152	55.3	50	17	US-10-885-482-8	Sequence 8, Appl
20	108	39.3	19	15	US-10-601-953-41	Sequence 41, Appl
21	107	38.9	19	15	US-10-601-953-47	Sequence 47, Appl
22	99	36.0	17	15	US-10-601-953-46	Sequence 46, Appl
23	96	34.9	17	15	US-10-601-953-40	Sequence 40, Appl
24	88	32.0	15	15	US-10-601-953-45	Sequence 45, Appl
25	82	29.8	15	15	US-10-601-953-39	Sequence 39, Appl
26	79	28.7	13	15	US-10-601-953-44	Sequence 44, Appl
27	68	24.7	13	15	US-10-601-953-38	Sequence 38, Appl
28	61.5	22.4	165	15	US-10-424-599-209220	Sequence 209220,
29	61	22.2	10	10	US-09-778-026-17	Sequence 17, Appl
30	61	22.2	10	14	US-10-119-537-12	Sequence 12, Appl
31	61	22.2	10	15	US-10-462-452-738	Sequence 738, App
32	61	22.2	10	15	US-10-601-953-843	Sequence 843, App
33	61	22.2	10	16	US-10-322-266-739	Sequence 739, App
34	61	22.2	10	16	US-10-648-854-17	Sequence 17, Appl
35	61	22.2	10	17	US-10-885-482-2	Sequence 2, Appl
36	59	21.5	11	15	US-10-601-953-37	Sequence 37, Appl
37	59	21.5	99	16	US-10-437-963-183064	Sequence 183064,
38	58	21.1	10	15	US-10-462-452-731	Sequence 731, App
39	58	21.1	10	15	US-10-462-452-732	Sequence 732, App
40	58	21.1	10	15	US-10-601-953-836	Sequence 836, App
41	58	21.1	10	15	US-10-601-953-837	Sequence 837, App
42	58	21.1	10	16	US-10-322-266-732	Sequence 732, App
43	58	21.1	10	16	US-10-322-266-733	Sequence 733, App
44	58	21.1	953	16	US-10-437-963-126195	Sequence 126195,
45	57.5	20.9	378	13	US-10-108-605-123	Sequence 123, App

ALIGNMENTS

RESULT 1
US-09-891-064A-2
; Sequence 2, Application US/09891064A
; Patent No. US20020082391A1
; GENERAL INFORMATION:
; APPLICANT: James M. Anderson
; APPLICANT: Christina M. Van Itallie
; TITLE OF INVENTION: Human Occludin, Its Uses and Enhancement of Drug
; TITLE OF INVENTION: Absorption Using Occludin Inhibitors
; FILE REFERENCE: OCR-754.CIP
; CURRENT APPLICATION NUMBER: US/09/891,064A
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 09/142,732
; PRIOR FILING DATE: 1998-09-15
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: MS DOS
; SEQ ID NO 2
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: peptide
; LOCATION: complete sequence
; OTHER INFORMATION: human occludin
US-09-891-064A-2

Query Match 100.0%; Score 275; DB 9; Length 522;
Best Local Similarity 100.0%; Pred.No. 3.9e-28;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GWNPTAQSSGLSGSIYALCNQFYTPAATGLYVDQYLHYCVVDPOQAIA 51
Db 196 GWNPTAQSSGLSGSIYALCNQFYTPAATGLYVDQYLHYCVVDPOQAIA 246

RESULT 2

```
US-10-176-847-104
; Sequence 104, Application US/10176847
; Publication No. US2003006836A1
; GENERAL INFORMATION:
; APPLICANT: Veiby, Petter Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
; FILE REFERENCE: MRI-039
; CURRENT APPLICATION NUMBER: US/10/176,847
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 104
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-176-847-104

Query Match      100.0%; Score 275; DB 14; Length 522;
Best Local Similarity 100.0%; Pred. No. 3.9e-28;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOBAIA 51
|
Db 196 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOBAIA 246
|

RESULT 3
US-10-601-953-900
; Sequence 900, Application US/10601953
; Publication No. US20040077540A1
; GENERAL INFORMATION:
; APPLICANT: Quay, Steven C.
; TITLE OF INVENTION: Compositions And Methods For Modulating Physiology Of Epithelial
; TITLE OF INVENTION: Junctional Adhesion Molecules For Enhanced Mucosal Delivery Of
; FILE REFERENCE: 02-03US
; CURRENT APPLICATION NUMBER: US/10/601,953
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 60/392,512
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 900
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 900
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
US-10-601-953-900

Query Match      100.0%; Score 275; DB 15; Length 522;
Best Local Similarity 100.0%; Pred. No. 3.9e-28;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOBAIA 51
|
Db 196 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOBAIA 246
|

RESULT 4
US-09-769-145-78
; Sequence 78, Application US/09769145
; Patent No. US20020168761A1
; GENERAL INFORMATION:
; APPLICANT: Gour, Barbara J.
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Ali, Anmar
; APPLICANT: Ni, Feng
; APPLICANT: Chen, Zhigang
; APPLICANT: Michaud, Stephanie
; APPLICANT: Wang, Shoameng
US-09-769-145-78

Query Match      95.6%; Score 263; DB 14; Length 48;
Best Local Similarity 100.0%; Pred. No. 9.2e-28;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48
|
Db 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48
|

US-10-105-008-78
; Sequence 78, Application US/10105008
; Publication No. US20030065136A1
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; APPLICANT: Farookhi, Riaz
; APPLICANT: Ali, Anmar
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING CELL ADHESION
; FILE REFERENCE: 100086.401D1
; CURRENT APPLICATION NUMBER: US/10/105,008
; CURRENT FILING DATE: 2002-03-22
; NUMBER OF SEQ ID NOS: 99
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 78
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Occludin cell
; OTHER INFORMATION: adhesion recognition sequence and flanking amino
; OTHER INFORMATION: acids
US-10-105-008-78

Query Match      95.6%; Score 263; DB 9; Length 48;
Best Local Similarity 100.0%; Pred. No. 9.2e-28;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48
|
Db 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48
|

US-10-105-008-78
; Sequence 78, Application US/10105008
; Publication No. US20030065136A1
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; APPLICANT: Farookhi, Riaz
; APPLICANT: Ali, Anmar
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING CELL ADHESION
; FILE REFERENCE: 100086.401D1
; CURRENT APPLICATION NUMBER: US/10/105,008
; CURRENT FILING DATE: 2002-03-22
; NUMBER OF SEQ ID NOS: 99
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 78
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Occludin cell
; OTHER INFORMATION: adhesion recognition sequence and flanking amino
; OTHER INFORMATION: acids
US-10-105-008-78

Query Match      95.6%; Score 263; DB 14; Length 48;
Best Local Similarity 100.0%; Pred. No. 9.2e-28;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48
|
Db 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48
|

RESULT 6
US-10-006-869-62
; Sequence 62, Application US/10006869
; Publication No. US20030082166A1
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Symonds, James Matthew
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL
; TITLE OF INVENTION: CADHERIN-MEDIATED FUNCTIONS
; FILE REFERENCE: 100086.407C7
US-10-006-869-62

Query Match      95.6%; Score 263; DB 14; Length 48;
Best Local Similarity 100.0%; Pred. No. 9.2e-28;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48
|
Db 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48
|
```

```

; CURRENT APPLICATION NUMBER: US/10/006,869
; CURRENT FILING DATE: 2001-12-03
; NUMBER OF SEQ ID NOS: 4052
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Occludin Cell
; OTHER INFORMATION: Adhesion Recognition Sequence
US-10-006-869-62

Query Match          95.6%; Score 263; DB 14; Length 48;
Best Local Similarity 100.0%; Pred. No. 9.2e-28;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVNPTAQSSGLSGSIYALCNQFYTPAATGLYVDQYLYHYCVDPQE 48
Db 1 GVNPTAQSSGLSGSIYALCNQFYTPAATGLYVDQYLYHYCVDPQE 48

RESULT 7
US-10-193-653-69
; Sequence 69, Application US/10193653
; Publication No. US20030109454A1
; GENERAL INFORMATION:
; APPLICANT: Doherty, Patrick
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING ADHESION MOLECULE
; FILE REFERENCE: 100086.403C2
; CURRENT APPLICATION NUMBER: US/10/193,653
; CURRENT FILING DATE: 2002-07-10
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 69
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Solid Phase
; OTHER INFORMATION: Synthesis
US-10-193-653-69

Query Match          95.6%; Score 263; DB 14; Length 48;
Best Local Similarity 100.0%; Pred. No. 9.2e-28;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVNPTAQSSGLSGSIYALCNQFYTPAATGLYVDQYLYHYCVDPQE 48
Db 1 GVNPTAQSSGLSGSIYALCNQFYTPAATGLYVDQYLYHYCVDPQE 48

RESULT 8
US-10-395-032-62
; Sequence 62, Application US/10395032
; Publication No. US20030229199A1
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Symonds, James Matthew
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL
; TITLE OF INVENTION: CADHERIN-MEDIATED FUNCTIONS
; FILE REFERENCE: 100086.407C9
; CURRENT APPLICATION NUMBER: US/10/395,032
; CURRENT FILING DATE: 2003-03-21
; NUMBER OF SEQ ID NOS: 4052
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 48
; TYPE: PRT

```

```

; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Occludin Cell
; OTHER INFORMATION: Adhesion Recognition Sequence
US-10-395-032-62

Query Match          95.6%; Score 263; DB 15; Length 48;
Best Local Similarity 100.0%; Pred. No. 9.2e-28;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVNPTAQSSGLSGSIYALCNQFYTPAATGLYVDQYLYHYCVDPQE 48
Db 1 GVNPTAQSSGLSGSIYALCNQFYTPAATGLYVDQYLYHYCVDPQE 48

RESULT 9
US-10-425-557-78
; Sequence 78, Application US/10425557
; Publication No. US20040006011A1
; GENERAL INFORMATION:
; APPLICANT: Gour, Barbara J.
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Ali, Ammar
; APPLICANT: Ni, Feng
; APPLICANT: Chen, Zhigang
; APPLICANT: Michaud, Stephanie Denise
; APPLICANT: Wang, Shaomeng
; APPLICANT: Hu, Zengjian
; TITLE OF INVENTION: PEPTIDOMIMETIC MODULATORS OF CELL ADHESION
; FILE REFERENCE: 100086.401C16
; CURRENT APPLICATION NUMBER: US/10/425,557
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 78
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Occludin cell
; OTHER INFORMATION: adhesion recognition sequence and flanking amino
; OTHER INFORMATION: acids
US-10-425-557-78

Query Match          95.6%; Score 263; DB 15; Length 48;
Best Local Similarity 100.0%; Pred. No. 9.2e-28;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVNPTAQSSGLSGSIYALCNQFYTPAATGLYVDQYLYHYCVDPQE 48
Db 1 GVNPTAQSSGLSGSIYALCNQFYTPAATGLYVDQYLYHYCVDPQE 48

RESULT 10
US-10-412-701-78
; Sequence 78, Application US/10412701
; Publication No. US20040058864A1
; GENERAL INFORMATION:
; APPLICANT: Gour, Barbara J.
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Ali, Ammar
; APPLICANT: Ni, Feng
; APPLICANT: Chen, Zhigang
; APPLICANT: Michaud, Stephanie Denise
; APPLICANT: Wang, Shaomeng
; APPLICANT: Hu, Zengjian
; TITLE OF INVENTION: PEPTIDOMIMETIC MODULATORS OF CELL ADHESION
; FILE REFERENCE: 100086.413C2
; CURRENT APPLICATION NUMBER: US/10/412,701
; CURRENT FILING DATE: 2003-04-19
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 78

```

```
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Occludin cell
; OTHER INFORMATION: adhesion recognition sequence and flanking amino
; OTHER INFORMATION: acids
US-10-412-701-78

Query Match          95.6%; Score 263; DB 15; Length 48;
Best Local Similarity 100.0%; Pred. No. 9.2e-28;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVNPTAQSSGSLYGSQIYALCNQFVTPAATGLYVDQYLYHYCVDPQE 48
Db 1 GVNPTAQSSGSLYGSQIYALCNQFVTPAATGLYVDQYLYHYCVDPQE 48

RESULT 11
US-10-632-678-78
; Sequence 78, Application US/10632678
; Publication No. US20040106545A1
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; APPLICANT: Farookhi, Riaz
; APPLICANT: Ali, Ammar
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING CELL ADHESION
; FILE REFERENCE: 100086.401C18
; CURRENT APPLICATION NUMBER: US/10/632,678
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 78
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Occludin cell
; OTHER INFORMATION: adhesion recognition sequence and flanking amino
; OTHER INFORMATION: acids
US-10-632-678-78

Query Match          95.6%; Score 263; DB 16; Length 48;
Best Local Similarity 100.0%; Pred. No. 9.2e-28;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVNPTAQSSGSLYGSQIYALCNQFVTPAATGLYVDQYLYHYCVDPQE 48
Db 1 GVNPTAQSSGSLYGSQIYALCNQFVTPAATGLYVDQYLYHYCVDPQE 48

RESULT 12
US-10-714-564A-1352
; Sequence 1352, Application US/10714564A
; Publication No. US20040175361A1
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Michaud, Stephanie D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING
; TITLE OF INVENTION: FUNCTIONS OF NONCLASSICAL CADHERINS
; FILE REFERENCE: 100086.418
; CURRENT APPLICATION NUMBER: US/10/714,564A
; CURRENT FILING DATE: 2003-11-14
; NUMBER OF SEQ ID NOS: 1402
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1352
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Occludin CAR sequence
US-10-714-564A-1352
```

```
Query Match          95.6%; Score 263; DB 16; Length 48;
Best Local Similarity 100.0%; Pred. No. 9.2e-28;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVNPTAQSSGSLYGSQIYALCNQFVTPAATGLYVDQYLYHYCVDPQE 48
Db 1 GVNPTAQSSGSLYGSQIYALCNQFVTPAATGLYVDQYLYHYCVDPQE 48

RESULT 13
US-10-885-482-5
; Sequence 5, Application US/10885482
; Publication No. US20050037973A1
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Symonds, James Matthew
; APPLICANT: Gour, Barbara J.
; APPLICANT: Alexander, J. Steven
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR CANCER THERAPY
; FILE REFERENCE: 100086.405CS
; CURRENT APPLICATION NUMBER: US/10/885,482
; CURRENT FILING DATE: 2004-07-06
; PRIOR APPLICATION NUMBER: 09/450,073
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: 09/222,373
; PRIOR FILING DATE: 1998-12-29
; PRIOR APPLICATION NUMBER: 09/001,511
; PRIOR FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-885-482-5

Query Match          95.6%; Score 263; DB 17; Length 48;
Best Local Similarity 100.0%; Pred. No. 9.2e-28;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVNPTAQSSGSLYGSQIYALCNQFVTPAATGLYVDQYLYHYCVDPQE 48
Db 1 GVNPTAQSSGSLYGSQIYALCNQFVTPAATGLYVDQYLYHYCVDPQE 48

RESULT 14
US-09-778-026-15
; Sequence 15, Application US/09778026
; Publication No. US20030013655A1
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR REGULATING
; CELL ADHESION
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/778,026
; FILING DATE: 05-Feb-2001
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
```


Search completed: April 1, 2005, 10:03:20
Job time : 137.553 secs

NAME: Maki, David J.
REGISTRATION NUMBER: 32,391
REFERENCE/DOCKET NUMBER: 100086.402
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 47 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-09-778-026-15

Query Match 91.3%; Score 251; DB 10; Length 47;
Best Local Similarity 97.9%; Pred. No. 3.7e-26;
Matches 46; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPQ 47
Db 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPQ 47

RESULT 15

US-10-648-854-15
Sequence 15, Application US/10648854
Publication No. US20040132651A1
GENERAL INFORMATION:
APPLICANT: Blaschuk, Orest W.
Gour, Barbara J.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR REGULATING
CELL ADHESION

NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed IP Law Group PLLC
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/10/648,854
APPLICATION NUMBER: US/10/648,854
FILING DATE: 25-Aug-2003
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Jeffrey E. Hundley
REGISTRATION NUMBER: 42,676
REFERENCE/DOCKET NUMBER: 100086.402C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 47 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-10-648-854-15

Query Match 91.3%; Score 251; DB 16; Length 47;
Best Local Similarity 97.9%; Pred. No. 3.7e-26;
Matches 46; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPQ 47
Db 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPQ 47

BEST AVAILABLE COPY

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2005, 09:16:13 ; Search time 6.73096 Seconds
(without alignments)
565.610 Million cell updates/sec

Title: US-09-891-064A-2_COPY_196_246

Perfect score: 275
Sequence: 1 GVNPTAQSSGLYSGIYAL.....LYVDQLYHYCVDPQEAIA 51

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/prodata/1/1aa/5A-COMB.pep:**

2: /cgn2_6/prodata/1/1aa/5B-COMB.pep:**

3: /cgn2_6/prodata/1/1aa/6A-COMB.pep:**

4: /cgn2_6/prodata/1/1aa/6B-COMB.pep:**

5: /cgn2_6/prodata/1/1aa/PCTUS-COMB.pep:**

6: /cgn2_6/prodata/1/1aa/backfiles1.pep:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	275	100.0	522	3	US-09-142-732-2
2	275	100.0	522	4	US-08-945-826-2
3	275	100.0	522	4	US-09-197-503-2
4	263	95.6	48	3	US-09-222-373-5
5	263	95.6	48	3	US-08-996-679-59
6	263	95.6	48	3	US-09-001-511-5
7	263	95.6	48	3	US-09-113-977C-69
8	263	95.6	48	3	US-09-510-616-5
9	263	95.6	48	3	US-09-248-074-78
10	263	95.6	48	3	US-09-187-859-62
11	263	95.6	48	4	US-09-458-870-78
12	263	95.6	48	4	US-09-351-048A-69
13	263	95.6	48	4	US-09-839-542B-62
14	263	95.6	48	4	US-09-450-073-5
15	263	95.6	48	4	US-10-193-653-69
16	255	92.7	521	4	US-08-945-826-6
17	255	92.7	521	4	US-09-197-503-6
18	251	91.3	47	3	US-08-939-853A-15
19	244	88.7	521	4	US-08-945-826-4
20	244	88.7	521	4	US-09-197-503-4
21	243	88.4	48	3	US-09-222-373-6
22	243	88.4	48	3	US-09-001-511-6
23	243	88.4	48	3	US-09-510-616-6
24	243	88.4	48	4	US-09-450-073-6
25	232	84.4	48	3	US-09-222-373-7
26	232	84.4	48	3	US-09-001-511-7
27	232	84.4	48	3	US-09-510-616-7

28	232	84.4	48	4	US-09-450-073-7	Sequence 7, Appli
29	221.5	80.5	51	3	US-09-222-373-46	Sequence 46, Appl
30	221.5	80.5	51	3	US-09-001-511-46	Sequence 46, Appl
31	221.5	80.5	51	3	US-09-510-616-46	Sequence 46, Appl
32	221.5	80.5	51	4	US-09-450-073-46	Sequence 46, Appl
33	152	55.3	50	3	US-09-222-373-8	Sequence 8, Appli
34	152	55.3	50	3	US-09-001-511-8	Sequence 8, Appli
35	152	55.3	50	3	US-09-510-616-8	Sequence 8, Appli
36	152	55.3	50	4	US-09-450-073-8	Sequence 8, Appli
37	62.5	22.7	224	4	US-09-270-767-60592	Sequence 60592, A
38	62.5	22.7	340	4	US-09-270-767-45097	Sequence 45097, A
39	61	22.2	10	3	US-09-222-373-2	Sequence 2, Appli
40	61	22.2	10	3	US-08-996-679-60	Sequence 60, Appl
41	61	22.2	10	3	US-08-939-853A-17	Sequence 17, Appl
42	61	22.2	10	3	US-09-001-511-2	Sequence 2, Appli
43	61	22.2	10	3	US-09-510-616-2	Sequence 2, Appli
44	61	22.2	10	3	US-09-324-541-12	Sequence 12, Appl
45	61	22.2	10	4	US-09-282-029A-478	Sequence 478, App

ALIGNMENTS

RESULT 1

US-09-142-732-2

; Sequence 2, Application US/09142732

; Patent No. 6252045

; GENERAL INFORMATION:

; APPLICANT: James M. Anderson

; APPLICANT: Christina M. Van Itallie

; TITLE OF INVENTION: Human Occludin, Its Uses and

; TITLE OF INVENTION: Enhancement of Drug Absorption Using Occludin Inhibitors

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Yale University Medical School

; ADDRESSEE: Section of Digestive Diseases

; ADDRESSEE: Department of Internal Medicine

; STREET: 333 Cedar Street, LCI 105

; CITY: New Haven

; STATE: Connecticut

; COUNTRY: United States of America

; ZIP: 06520-8057

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" 1.44 Mb diskette

; COMPUTER: IBM PC

; OPERATING SYSTEM: MS DOS

; SOFTWARE: Word Processing

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/142,732

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US97/05809

; FILING DATE: March 14, 1997

; APPLICATION NUMBER: U.S. 60/013,625

; FILING DATE: March 15, 1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Mary M. Krinsky

; REGISTRATION NUMBER: 32423

; REFERENCE/DOCKET NUMBER: 1751-P0016B.PCT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 203-324-6155

; TELEFAX: 203-327-1096

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 522

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: polypeptide

; FRAGMENT TYPE: complete sequence

; FEATURE:

```
; NAME/KEY: human occludin
US-09-142-732-2

Query Match      100.0%; Score 275; DB 3; Length 522;
Best Local Similarity 100.0%; Pred. No. 4.2e-29;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOEAIA 51
Db 196 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOEAIA 246

RESULT 2
US-08-945-826-2
; Sequence 2, Application US/08945826
; Patent No. 6489460
; GENERAL INFORMATION:
; APPLICANT: TSUKITA, Shoichiro
; TITLE OF INVENTION: HUMAN ADHESION MOLECULE OCCLUDIN
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/945,826
; FILING DATE: 05-NOV-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C. Joseph
; REGISTRATION NUMBER: 32,350
; REFERENCE/DOCKET NUMBER: 0425-0660P
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 522 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-945-826-2

Query Match      100.0%; Score 275; DB 4; Length 522;
Best Local Similarity 100.0%; Pred. No. 4.2e-29;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOEAIA 51
Db 196 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOEAIA 246

RESULT 3
US-09-197-503-2
; Sequence 2, Application US/09197503
; Patent No. 655286
; GENERAL INFORMATION:
; APPLICANT: TSUKITA, Shoichiro
; TITLE OF INVENTION: HUMAN ADHESION MOLECULE OCCLUDIN
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/197,503
; FILING DATE: 05-NOV-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C. Joseph
; REGISTRATION NUMBER: 32,350
; REFERENCE/DOCKET NUMBER: 0425-0660P
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 522 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-945-826-2

Query Match      100.0%; Score 275; DB 4; Length 522;
Best Local Similarity 100.0%; Pred. No. 4.2e-29;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOEAIA 51
Db 196 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOEAIA 246

RESULT 4
US-09-222-373-5
; Sequence 5, Application US/09222373
; Patent No. 6110747
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING TISSUE
; FILE REFERENCE: 10086.405C1
; CURRENT APPLICATION NUMBER: US/09/222,373
; CURRENT FILING DATE: 1998-12-29
; EARLIER APPLICATION NUMBER: US 09/001,511
; EARLIER FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-222-373-5

Query Match      95.6%; Score 263; DB 3; Length 48;
Best Local Similarity 100.0%; Pred. No. 8.8e-29;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48
Db 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48

RESULT 5
US-08-996-679-59
; Sequence 59, Application US/08996679
; Patent No. 6169071
```

```

; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING
; TITLE OF INVENTION: CELL ADHESION
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/996,679
; FILING DATE: 23-DEC-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 100086.401C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 622-6031
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48 amino acids
; TYPE: amino acid
; STRANDEDNESS: linear
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-996-679-59

```

```

Query Match 95.6%; Score 263; DB 3; Length 48;
Best Local Similarity 100.0%; Pred. No. 8.8e-29;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVNPTAQSSGLSGSIYALCNQFYTPAATGLYVDQYLHYCVVDPOE 48
Db 1 GVNPTAQSSGLSGSIYALCNQFYTPAATGLYVDQYLHYCVVDPOE 48

RESULT 6
US-09-001-511-5
; Sequence 5, Application US/09001511
; Patent No. 6248864
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING
; TITLE OF INVENTION: TISSUE PERMEABILITY
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/001,511
; FILING DATE: 31-DEC-1997
; CLASSIFICATION: 530

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 100086.405
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-001-511-5

Query Match 95.6%; Score 263; DB 3; Length 48;
Best Local Similarity 100.0%; Pred. No. 8.8e-29;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVNPTAQSSGLSGSIYALCNQFYTPAATGLYVDQYLHYCVVDPOE 48
Db 1 GVNPTAQSSGLSGSIYALCNQFYTPAATGLYVDQYLHYCVVDPOE 48

RESULT 7
US-09-113-977C-69
; Sequence 69, Application US/09113977C
; Patent No. 6277824
; GENERAL INFORMATION:
; APPLICANT: Doherty, Patrick
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING ADHESION MOLECULE
; TITLE OF INVENTION: FUNCTION
; FILE REFERENCE: 100086.403
; CURRENT APPLICATION NUMBER: US/09/113,977C
; CURRENT FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 69
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Solid Phase
; OTHER INFORMATION: Synthesis
; US-09-113-977C-69

Query Match 95.6%; Score 263; DB 3; Length 48;
Best Local Similarity 100.0%; Pred. No. 8.8e-29;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVNPTAQSSGLSGSIYALCNQFYTPAATGLYVDQYLHYCVVDPOE 48
Db 1 GVNPTAQSSGLSGSIYALCNQFYTPAATGLYVDQYLHYCVVDPOE 48

RESULT 8
US-09-510-616-5
; Sequence 5, Application US/09510616
; Patent No. 6310177
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Symonds, James Matthew
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING
; TISSUE PERMEABILITY
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED IP LAW GROUP PLLC
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington

```

;; COUNTRY: USA
;; ZIP: 98104
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/510,616
;; FILING DATE: 22-Feb-2000
;; CLASSIFICATION: <Unknown>
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Christiansen, William T.
;; REGISTRATION NUMBER: 44,614
;; REFERENCE/DOCKET NUMBER: 100086.405C3
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (206)622-4900
;; TELEFAX: (206) 682-6031
;; INFORMATION FOR SEQ ID NO: 5:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 48 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-510-616-5

Query Match 95.6%; Score 263; DB 3; Length 48;
Best Local Similarity 100.0%; Pred. No. 8.8e-29;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48
Db 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48

RESULT 9

US-09-248-074-78
; Sequence 78, Application US/09248074
; Patent No. 6346512
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING CELL ADHESION
; FILE REFERENCE: 100086.401C5
; CURRENT APPLICATION NUMBER: US/09/248,074
; CURRENT FILING DATE: 1999-02-10
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 78
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Occludin cell
; OTHER INFORMATION: adhesion recognition sequence and flanking amino
; OTHER INFORMATION: acids
US-09-248-074-78

Query Match 95.6%; Score 263; DB 3; Length 48;
Best Local Similarity 100.0%; Pred. No. 8.8e-29;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48
Db 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48

RESULT 10

US-09-187-859-62
; Sequence 62, Application US/09187859A
; Patent No. 6358920
; GENERAL INFORMATION:

;; APPLICANT: Blaschuk, Orest W.
;; APPLICANT: Gour, Barbara J.
;; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL
;; TITLE OF INVENTION: CADHERIN-MEDIATED FUNCTIONS
;; FILE REFERENCE: 100086.407C1
;; CURRENT APPLICATION NUMBER: US/09/187,859A
;; CURRENT FILING DATE: 1998-11-06
;; NUMBER OF SEQ ID NOS: 4052
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 62
;; LENGTH: 48
;; TYPE: PRT
;; ORGANISM: Unknown
;; FEATURE:
;; OTHER INFORMATION: Description of Unknown Organism: Occludin Cell
;; OTHER INFORMATION: Adhesion Recognition Sequence
US-09-187-859-62

Query Match 95.6%; Score 263; DB 3; Length 48;
Best Local Similarity 100.0%; Pred. No. 8.8e-29;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48
Db 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48

RESULT 11

US-09-458-870-78
; Sequence 78, Application US/09458870
; Patent No. 6465427
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; APPLICANT: Farookhi, Riaz
; APPLICANT: Ali, Aumaz
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING CELL ADHESION
; FILE REFERENCE: 100086.401C8
; CURRENT APPLICATION NUMBER: US/09/458,870
; CURRENT FILING DATE: 1999-12-10
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 78
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Occludin cell
; OTHER INFORMATION: adhesion recognition sequence and flanking amino
; OTHER INFORMATION: acids
US-09-458-870-78

Query Match 95.6%; Score 263; DB 4; Length 48;
Best Local Similarity 100.0%; Pred. No. 8.8e-29;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48
Db 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48

RESULT 12

US-09-351-048A-69
; Sequence 69, Application US/09351048A
; Patent No. 6472368
; GENERAL INFORMATION:
; APPLICANT: Doherty, Patrick
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING ADHESION MOLECULE
; TITLE OF INVENTION: FUNCTION
; FILE REFERENCE: 100086.403C1
; CURRENT APPLICATION NUMBER: US/09/351,048A

```
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 69
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Solid Phase
; OTHER INFORMATION: Synthesis
US-09-351-048A-69

Query Match          95.6%; Score 263; DB 4; Length 48;
Best Local Similarity 100.0%; Pred. No. 8.8e-29;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVNPTAQSSGLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48
Db 1 GVNPTAQSSGLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48

RESULT 13
US-09-839-542B-62
; Sequence 62, Application US/09839542B
; Patent No. 6569996
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Symonds, James Matthew
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL
; FILE OF INVENTION: CADHERIN-MEDIATED FUNCTIONS
; FILE REFERENCE: 100086.407D1
; CURRENT APPLICATION NUMBER: US/09/839,542B
; CURRENT FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 4052
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Occludin Cell
; OTHER INFORMATION: Adhesion Recognition Sequence
US-09-839-542B-62

Query Match          95.6%; Score 263; DB 4; Length 48;
Best Local Similarity 100.0%; Pred. No. 8.8e-29;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVNPTAQSSGLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48
Db 1 GVNPTAQSSGLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48

RESULT 14
US-09-450-073-5
; Sequence 5, Application US/09450073
; Patent No. 6757807
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Symonds, James Matthew
; APPLICANT: Gour, Barbara J.
; APPLICANT: Alexander, J. Steven
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR CANCER THERAPY
; FILE REFERENCE: 100086.405C2
; CURRENT APPLICATION NUMBER: US/09/450,073
; CURRENT FILING DATE: 1999-11-29
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Homo sapiens
```

US-09-450-073-5

```
Query Match          95.6%; Score 263; DB 4; Length 48;
Best Local Similarity 100.0%; Pred. No. 8.8e-29;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVNPTAQSSGLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48
Db 1 GVNPTAQSSGLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48

RESULT 15
US-10-193-653-69
; Sequence 69, Application US/10193653
; Patent No. 6806255
; GENERAL INFORMATION:
; APPLICANT: Doherty, Patrick
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING ADHESION MOLECULE
; FILE OF INVENTION: FUNCTION
; FILE REFERENCE: 100086.403C2
; CURRENT APPLICATION NUMBER: US/10/193,653
; CURRENT FILING DATE: 2002-07-10
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 69
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Solid Phase
; OTHER INFORMATION: Synthesis
US-10-193-653-69

Query Match          95.6%; Score 263; DB 4; Length 48;
Best Local Similarity 100.0%; Pred. No. 8.8e-29;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVNPTAQSSGLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48
Db 1 GVNPTAQSSGLYGSQIYALCNQFYTPAATGLYVDQYLYHYCVVDPOE 48

Search completed: April 1, 2005, 10:04:45
Job time : 7.73096 secs
```

BEST AVAILABLE COPY

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2005, 09:36:47 ; Search time 5.43655 Seconds
(without alignments)
902.605 Million cell updates/sec

Title: US-09-891-064A-2_COPY_196_246
Perfect score: 275
Sequence: 1 GVNPTAQSSGSLYGSQIYAL.....LYVDQYLHYCVDPQEAIA 51

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	275	100.0	522	2 G02533	occludin - human
2	144	52.4	504	2 A49467	occludin - chicken
3	60	21.8	533	2 D87556	hypothetical prote
4	58	21.1	100	1 CCQF2M	cytochrome c2, iso
5	58	21.1	294	2 AD1795	xylose operon regu
6	57.5	20.9	553	2 A72228	hypothetical prote
7	57	20.7	379	2 A72409	mannosyl transfera
8	57	20.7	579	2 AC2200	potassium-transport
9	56.5	20.5	256	1 JQ1886	coat protein - tom
10	56	20.4	57	1 INLMS	insulin - sea lamp
11	56	20.4	223	1 B48640	regulatory protein
12	56	20.4	223	2 AB2354	nitrogen-responsiv
13	56	20.4	331	2 AF1349	tryptophanyl-tRNA
14	56	20.4	331	2 A11719	cryptophanyl-tRNA
15	56	20.4	542	2 T46464	hypothetical prote
16	55.5	20.2	987	2 D97029	ribonucleotide red
17	55	20.0	225	1 A48640	global nitrogen re
18	55	20.0	330	2 B2415	probable serine pr
19	55	20.0	464	1 VCXFTI	major capsid prote
20	55	20.0	472	1 VCXFSI	major capsid prote
21	55	20.0	557	2 A57050	K-glypican precurs
22	55	20.0	867	2 F81721	polymorphic membra
23	55	20.0	947	2 H84866	hypothetical prote
24	54.5	19.8	303	2 T42699	hypothetical prote
25	54.5	19.8	2825	2 T14271	DocA protein, stre
26	54	19.6	324	2 S62543	hypothetical prote
27	54	19.6	378	2 S76183	hypothetical prote
28	54	19.6	588	2 T26193	hypothetical prote
29	54	19.6	590	2 T06626	hypothetical prote

30	54	19.6	799	2 F83456	xanthine dehydroge
31	54	19.6	1403	2 S64142	hypothetical prote
32	53.5	19.5	613	2 B90294	hypothetical prote
33	53	19.3	304	2 G90839	L, D-carboxypptid
34	53	19.3	304	2 G85697	L, D-carboxypptid
35	53	19.3	304	2 G85697	Muramoyltetrapepti
36	53	19.3	391	2 B70608	probable PPE prote
37	53	19.3	396	2 H70741	translation releas
38	53	19.3	437	1 A48061	translation releas
39	53	19.3	437	1 S50853	putrescine/ornithi
40	53	19.3	455	2 T44519	hypothetical prote
41	53	19.3	497	2 T29791	probable metabolit
42	53	19.3	543	2 T39345	steroid/thyroid/re
43	53	19.3	583	2 T34121	ribonucleotide red
44	53	19.3	806	2 T41930	hypothetical prote
45	52.5	19.1	314	2 H90638	hypothetical prote

ALIGNMENTS

RESULT 1

G02533
occludin - human
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 09-Jul-2004
C:Accession: G02533
R:Van Itallie, C.M.
submitted to the EMBL Data Library, April 1996
A:Reference number: H01403
A:Accession: G02533
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-522 <V>
A:Cross-references: UNIPROT:Q16625; EMBL:U53823; NID:G13222281; PIDN:AAB00195.1; PID:G13222281
C:Superfamily: occludin

Query Match 100.0%; Score 275; DB 2; Length 522;
Best Local Similarity 100.0%; Pred. No. 1.2e-26;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLHYCVDPQEAIA 51
|||||
DB 196 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLHYCVDPQEAIA 246
|||||

RESULT 2

A49467
occludin - chicken
C:Species: Gallus gallus (chicken)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004
R:Furuse, M.; Hirase, T.; Itoh, M.; Nagafuchi, A.; Yonemura, S.; Tsukita, S.; Tsukita, E.; J. Cell Biol. 123, 1777-1788, 1993
A:Title: Occludin: a novel integral membrane protein localizing at tight junctions.
A:Reference number: A49467; MUID:94103332; PMID:8276896
A:Accession: A49467
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-504 <FUR>
A:Cross-references: UNIPROT:Q91049; GB:D21837; NID:G464148; PIDN:BAA04865.1; PID:G540494
C:Superfamily: occludin
C:Keywords: liver; membrane protein

Query Match 52.4%; Score 144; DB 2; Length 504;
Best Local Similarity 52.9%; Pred. No. 3.3e-10;
Matches 27; Conservative 7; Mismatches 13; Indels 4; Gaps 1;

QY 1 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLHYCVDPQEAIA 51
|||||
DB 184 GVNPTAQSSGSLYGSQIYALCNQFYTPAATGLYVDQYLHYCVDPQEAIA 230
|||||

```
RESULT 3
D87556
hypothetical protein CC2477 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C:Accession: D87556
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Emolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A>Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: D87556
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-533 <STO>
A:CROSS-references: UNIPROT:Q9A5H0; GB:AE005673; NID:gi3424028; PIDN:AAK24448.1; GSPDB:G
C:Genetics:
A:Gene: CC2477

Query Match 21.8%; Score 60; DB 2; Length 533;
Best Local Similarity 32.7%; Pred. No. 12;
Matches 17; Conservative 7; Mismatches 24; Indels 4; Gaps 1;

QY 4 PTAQSSGLYGSQIYALCNQFVTPAATGLY----VDQYLHYCVVDPOEATA 51
Db 213 PFGEPSDLQAEAGKPSQCPAPALDLFRNDVDVLYNVLAVQTGDVSA 264

RESULT 4
CCQF2M
cytochrome c2, iso-1 - Rhodospirillum molischianum
C:Species: Rhodospirillum molischianum
C>Date: 24-Apr-1984 #sequence_revision 24-Apr-1984 #text_change 09-Jul-2004
C:Accession: A00079
R:Amblar, R.P.; Meyer, T.E.; Bartsch, R.G.; Kamen, M.D.
unpublished results, cited by Amblar, R.P., in Evolution of Protein Molecules, Matsubara
an, Tokyo, 1978
A:Reference number: A94458
A:Accession: A00079
A:Molecule type: protein
A:Residues: 1-100 <AMB>
A:CROSS-references: UNIPROT:P00087
C:Superfamily: cytochrome c; cytochrome c homology
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; photosynthesis
F:1-96/Domain: cytochrome c homology <CYC>
F:11.14/Binding site: heme (Cys) (covalent) #status predicted
F:15.76/Binding site: heme iron (His, Met) (axial ligands) #status predicted

Query Match 21.1%; Score 58; DB 1; Length 100;
Best Local Similarity 28.8%; Pred. No. 3.7;
Matches 15; Conservative 10; Mismatches 25; Indels 2; Gaps 1;

QY 1 GVNPTAQSSGLYGSQIYALCNQFVTPA--ATGLYVDQYLHYCVVDPOEAI 50
Db 20 GNGVGPSPSLSGYGRKVGAPNPKYSPHLAGSGMTIDDLTKYLANPKETI 71

RESULT 5
AD1795
xylose operon regulatory protein and to glucose kinase homolog lin2907 [imported] - List
C:Species: Listeria innocua
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AD1795
R:Glaeser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Smoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A>Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AD1795
```

```
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-294 <GLA>
A:CROSS-references: UNIPROT:Q926Y3; GB:AL592022; PIDN:CAC98132.1; PID:gl6415448; GSPDB:G
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: lin2907
C:Superfamily: glucose kinase; glucose kinase homology

Query Match 21.1%; Score 58; DB 2; Length 294;
Best Local Similarity 33.9%; Pred. No. 11;
Matches 19; Conservative 6; Mismatches 11; Indels 20; Gaps 3;

QY 6 AOSSG----SLYGSQIYA-----LCNQFVTPAATGLYVDQYLHYCVVD 46
Db 185 AOLGRPLKEITGEIFANYDAHDPISERLINEFYTGICTGLYNLIYLF-----DP 235

RESULT 6
A72228
hypothetical protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: A72228
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: A72228
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-553 <ARN>
A:CROSS-references: UNIPROT:Q9X1Y5; GB:AE001807; GB:AE000512; NID:g4982216; PIDN:AAD3671
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1652

Query Match 20.9%; Score 57.5; DB 2; Length 553;
Best Local Similarity 34.3%; Pred. No. 26;
Matches 12; Conservative 6; Mismatches 16; Indels 1; Gaps 1;

QY 11 SLYGSQIYALCNQFVTPAATGLYVDQYLHYCVVD 45
Db 31 TVFSGEIIYVCAEDLAPAGVG-YIHSNGYHIIVYD 64

RESULT 7
AF2409
mannosyl transferase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A>Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: AF2409
R:Kanezo, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yanada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A>Title: Complete Genomic
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AF2409
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-379 <KUR>
A:CROSS-references: UNIPROT:Q8YMU7; GB:BA000019; PIDN:BAB76529.1; PID:gl7133967; GSPDB:G
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all4830
C:Superfamily: hypothetical protein sll1534

Query Match 20.7%; Score 57; DB 2; Length 379;
Best Local Similarity 38.9%; Pred. No. 20;
Matches 14; Conservative 7; Mismatches 7; Indels 8; Gaps 2;
```

RESULT 12

AH2354
nitrogen-responsive regulatory protein ntca [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:accession: AH2354
R:Kaneoko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigun
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840

```

A;Gene: trpS
C;Superfamily: tryptophan-tRNA ligase

Query Match          20.4%; Score 56; DB 2; Length 331;
Best Local Similarity 32.1%; Pred. No. 23;
Matches 18; Conservative 9; Mismatches 13; Indels 16; Gaps 4;

QY 1 GVNPTAQSGLSYGSGIYALCNQFYTPAATGLYVDQYLHYHCVD-----PQEI 50
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 8 GIQPSQLT---LGNYGAL-KQF-----GQDEYECFYCIVDSEHAITVPQDRL 53

RESULT 15
T46464
hypothetical protein DKFZp434P072.1 - human
C;Species: Homo sapiens (man)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C;Accession: T46464
R;Ansoorge, W.; Wirkner, U.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23028
A;Accession: T46464
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-542 <AA>
A;Cross-references: UNIPROT:Q9NTB1; EMBL:AL137421
A;Experimental source: adult testis; clone DKFZp434P072
C;Genetics:
A;Note: DKFZp434P072.1

Query Match          20.4%; Score 56; DB 2; Length 542;
Best Local Similarity 37.0%; Pred. No. 39;
Matches 17; Conservative 5; Mismatches 16; Indels 8; Gaps 2;

QY 3 NPTAQSGLSY-----GSGIYALCNQFYTPAATGLYVD---QYLYH 40
      ||||| ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 166 SPTAQSYSQYPPVDVSTYQDSTGYYVDPQTGLYDPSQYYN 211

Search completed: April 1, 2005, 10:16:13
Job time : 8.43655 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2005, 09:16:14 ; Search time 261.168 Seconds
(without alignments)
725.636 Million cell updates/sec

Title: US-09-891-064A-2_COPY_33_522
Perfect score: 2590
Sequence: 1 MHVRPMLSQPAYSPYDEI.....LKSLSHIKVMGVDRQKT 490

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_18Dec04:*

- 1: Geneseqp1980a:*
- 2: Geneseqp1990a:*
- 3: Geneseqp2000a:*
- 4: Geneseqp2001a:*
- 5: Geneseqp2002a:*
- 6: Geneseqp2003a:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004a:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2590	100.0	522	2	Aaw36052 Human occ
2	2590	100.0	522	2	Aaw34638 Human occ
3	2590	100.0	522	3	Aab35731 Human occ
4	2590	100.0	522	6	Abj37076 Human bre
5	2590	100.0	522	7	Add46545 Human pro
6	2590	100.0	522	8	Adi47189 Human occ
7	2498	96.4	508	8	Abm84985 Human dia
8	2498	96.4	508	8	Abm84986 Human dia
9	2335.5	90.2	521	2	Aaw36054 Mouse occ
10	2330.5	90.0	505	2	Aaw36053 Canine oc
11	2323.5	89.7	523	7	Add46543 Rat Prote
12	566	21.9	108	2	Aaw76214 Human occ
13	291	11.2	561	6	Abf58441 Human NOV
14	287.5	11.1	558	7	Adi21141 Novel hum
15	285	11.0	91	4	Aag74111 Human col
16	280.5	10.8	594	5	Aau11279 Human rec
17	275.5	10.6	475	5	Aau11278 Human rec
18	263	10.2	48	2	Aay17159 Peptide S
19	263	10.2	48	3	Aay96005 Human occ
20	263	10.2	48	3	Aay78216 Occludin
21	263	10.2	48	3	Aay64642 Occludin
22	263	10.2	48	3	Aag65438 Occludin
23	263	10.2	48	6	Abu60314 Mammalian
24	263	10.2	48	7	Abw01277 Classical
25	263	10.2	48	7	Abw01277 Occludin

26	263	10.2	48	8	ADK13612	Adk13612	Cadherin-
27	263	10.2	48	8	ADO70598	Ado70598	Occludin
28	263	10.2	48	8	ADP07175	Adp07175	Cell adhe
29	263	10.2	48	8	ADR74398	Adr74398	Human occ
30	251	9.7	47	2	AAy09102	Aay09102	Cell adhe
31	243	9.4	48	3	AAy96006	Aay96006	Mouse occ
32	243	9.4	48	8	ADR74399	Adr74399	Mouse occ
33	232	9.0	48	3	AAy96007	Aay96007	Dog occlu
34	232	9.0	48	8	ADR74400	Adr74400	Dog occlu
35	229.5	8.9	621	2	AAW76213	Aaw76213	Human ELL
36	229.5	8.9	625	4	ABG19597	Abg19597	Novel hum
37	221.5	8.6	51	3	AAy96009	Aay96009	Mammalian
38	219	8.5	640	2	AAW76212	Aaw76212	Human ELL
39	219	8.5	677	3	AAAB57048	Aab57048	Human pro
40	165	6.4	397	4	AAAG89128	Aag89128	Human bec
41	165	6.4	397	5	ABP69416	Abp69416	Human pol
42	165	6.4	397	8	ADL83175	Adl83175	Human PRO
43	165	6.4	397	8	ADN06048	Adn06048	Antipsori
44	165	6.4	397	8	ADP56404	Adp56404	Human PRO
45	156.5	6.0	363	4	ABG14568	Abg14568	Novel hum

ALIGNMENTS

RESULT 1
AAW36052
ID AAW36052 standard; protein; 522 AA.
XX
AC AAW36052;
XX
DT 25-MAR-1998 (first entry)
XX
DE Human occludin protein sequence.
XX
KW Occludin; adhesion molecule; membrane tight junction;
KW occludin localisation; membrane; occludin expression;
KW blood-brain barrier disorder.
XX
OS Homo sapiens.
XX
PN WO9732982-A1.
XX
PD 12-SEP-1997.
XX
PF 05-MAR-1997; 97WO-JP000665.
XX
PR 07-MAR-1996; 96JP-00049880.
PR 12-DEC-1996; 96JP-00331944.
(EISA) EISAI CO LTD.
PI Tsukita S;
XX
WI 1997-470546/43.
N-PSDB; AAT97972.
DNA encoding human, dog and mouse occludin(s) - useful for screening for
substances influencing occluding expression.
Claim 1; Page 15-17; 36pp; Japanese.

The present sequence represents human occludin, a novel adhesion molecule which is a constituent protein of membrane tight junctions. Antibodies raised against the protein are used to assay for occludin in samples, and for the study of occludin localisation in membranes (e.g. by immunofluorescent cell-staining). The transformatants and antibodies are used for screening of substances which potentially influence occludin expression. Therapeutic polynucleotides derived from the DNA encoding the present sequence are used for treatment of disorders involving the blood-brain barrier. Primers that hybridise to the occludin DNA sequence can be used to detect occludin DNA by PCR

Db 332 NGRSYPPDLYKSPPLVPEVAQEIPLTSLVDDFRQPRYSNDNLETPTKAGKA 391
 QY 359 KTEQDHYETDVTGGSDELEEDWIREYPPITSDQOQLYKRNFDTLGLQYKSLQSL 418
 Db 392 KTDPHYETDVTGGSDELEEDWIREYPPITSDQOQLYKRNFDTLGLQYKSLQSL 451
 QY 419 DEINKSLDKELDDYRESEBEYMAAAYNRLKQVKGADYKSKNHCKQLKSLSHI 478
 Db 452 DEVNKLSDRLDELDDYRESEBEYMAAAYNRLKQVKGADYKSKNHCKQLKSLSHI 511
 QY 479 KKMVGDDYDRQKT 490
 Db 512 KKMVGDDYDRKT 523

RESULT 12
 AAW76214
 ID AAW76214 standard; protein; 108 AA.
 AC AAW76214;
 XX 26-NOV-1998 (first entry)
 DT Human occludin protein fragment.
 DE
 KW ELL2; RNA polymerase II elongation factor; human; neoplastic disorder;
 KW leukaemia; inhibitor; diagnosis; susceptibility; occludin.
 XX Homo sapiens.
 OS
 PN WO9837194-A1.
 XX
 PD 27-AUG-1998.
 XX
 PF 19-FEB-1998; 98WO-US003177.
 XX
 PR 19-FEB-1997; 97US-0038447P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.
 XX
 PI Duan DR, Shilatfard A, Conaway JW, Conaway RC;
 XX WPI; 1998-467557/40.
 XX
 PT ELL2 RNA polymerase II elongation factor - for the prevention and
 PT treatment of neoplastic disorders, e.g. leukaemia.
 XX
 PS Disclosure; Fig 3; 84pp; English.

CC This sequence represents a human occludin protein which is used in a
 CC method to isolate RNA polymerase II elongation factor ELL2 which can be
 CC used in the design of treatments for neoplastic disorders (especially
 CC leukaemias). Antagonists of the polypeptides or nucleic acids that
 CC inhibit the expression of the protein, can be used to inhibit ELL2
 CC activity in a patient. A method for diagnosing or determining a
 CC susceptibility to neoplastic disorders comprises assaying ELL2 gene
 CC expression level in mammalian cells or body fluid and comparing these
 CC levels with a standard, such that an increase or decrease in levels
 CC indicates an increased or decreased susceptibility. The gene expression
 CC level is assayed using the antibody, or by detecting ELL2 mRNA levels.
 CC The host cells of may be used to identify compounds which bind to an ELL2
 CC polypeptide
 XX
 SQ Sequence 108 AA;
 Query Match 21.9%; Score 566; DB 2; Length 108;
 Best Local Similarity 100.0%; Pred. No. 5.1e-41;
 Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 383 DWIREYPPITSDQOQLYKRNFDTLGLQYKSLQSLDEINKSLDKELDDYRESEBEY 442
 |||||

Db 1 DWIREYPPITSDQOQLYKRNFDTLGLQYKSLQSLDEINKSLDKELDDYRESEBEY 60
 QY 443 MAADEYNRLKQVKGADYKSKNHCKQLKSLSHIKKMVGDDYDRQKT 490
 Db 61 MAADEYNRLKQVKGADYKSKNHCKQLKSLSHIKKMVGDDYDRQKT 108
 RESULT 13
 ABR58441
 ID ABR58441 standard; protein; 561 AA.
 XX ABR58441;
 AC ABR58441;
 XX 07-JUL-2003 (first entry)
 DT Human NOV43a.
 DE
 XX Human; NOV; antidiabetic; anorectic; antibacterial; virucide;
 KW immunomodulator; cytostatic; nootropic; neuroprotective; dyslipidaemia;
 KW antiparkinsonian; antilipaeamic; gene therapy; metabolic disorder;
 KW diabetes; obesity; infection; cachexia; cancer; Parkinson's disease;
 KW neurodegenerative disorder; Alzheimer's disease; immune disorder;
 KW haematopoietic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO2003029423-A2.
 XX
 PD 10-APR-2003.
 XX
 PF 02-OCT-2002; 2002WO-US031358.
 XX
 PR 02-OCT-2001; 2001US-0326483P.
 PR 05-OCT-2001; 2001US-0327342P.
 PR 09-OCT-2001; 2001US-0327917P.
 PR 09-OCT-2001; 2001US-0328029P.
 PR 09-OCT-2001; 2001US-0328044P.
 PR 09-OCT-2001; 2001US-0328056P.
 PR 12-OCT-2001; 2001US-0328849P.
 PR 15-OCT-2001; 2001US-0329414P.
 PR 17-OCT-2001; 2001US-0330142P.
 PR 22-OCT-2001; 2001US-0341058P.
 PR 24-OCT-2001; 2001US-0339266P.
 PR 24-OCT-2001; 2001US-0343629P.
 PR 29-OCT-2001; 2001US-0349575P.
 PR 01-NOV-2001; 2001US-0346357P.
 PR 12-APR-2002; 2002US-0371972P.
 PR 12-APR-2002; 2002US-0371980P.
 PR 17-APR-2002; 2002US-0373261P.
 PR 19-APR-2002; 2002US-0373805P.
 PR 23-APR-2002; 2002US-0374738P.
 PR 16-MAY-2002; 2002US-0381101P.
 PR 17-MAY-2002; 2002US-0381635P.
 PR 29-MAY-2002; 2002US-0383830P.
 PR 01-OCT-2002; 2002US-00262839.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Alsobrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;
 PI Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;
 PI Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;
 PI Rothenberg ME, Shinkets RA, Smithson G, Spytek KA, Taupier RJ;
 PI Vernet CAM, Voss EZ, Zerhusen BD, Zhong M;
 XX WPI; 2003-381625/36.
 DR N-PSDB; ACC72153.
 XX
 PT NOVX polypeptides and nucleic acids useful for diagnosing, preventing or
 PT treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or
 PT dyslipidaemia, and in chromosome mapping, tissue typing or
 PT pharmacogenomics.
 XX
 PS Claim 1; Page 240; 487pp; English.

BEST AVAILABLE COPY

Result No.	Query			Description	
	Score	Match	Length	ID	
1	2590	100.0	522	9	US-09-891-064A-2
2	2590	100.0	522	14	US-10-176-847-104
3	2590	100.0	522	15	US-10-601-953-900
4	566	21.9	108	13	US-10-028-780-8
5	291	11.2	561	15	US-10-262-839-158
6	285	11.0	91	14	US-10-106-598-4885
7	263	10.2	48	9	US-09-769-145-78
8	263	10.2	48	14	US-10-105-008-78
9	263	10.2	48	14	US-10-006-869-62
10	263	10.2	48	14	US-10-193-553-69
11	263	10.2	48	15	US-10-395-032-62
12	263	10.2	48	15	US-10-425-557-78
13	263	10.2	48	15	US-10-412-701-78
					Sequence 2, Appli
					Sequence 104, Appl
					Sequence 900, Appl
					Sequence 8, Appli
					Sequence 158, Appl
					Sequence 4885, Ap
					Sequence 78, Appl
					Sequence 62, Appl
					Sequence 69, Appl
					Sequence 62, Appl
					Sequence 78, Appl
					Sequence 78, Appl

Db 393 TTQDHVETDYYTGGBSCDLEBEDWIREYPPITSDQORQLYKRNFTGLQEQYKLSQSELDE 452
QY 421 INKLSRLDKELDDYREESEYMAADEYNRLKQVKGSDYKSKNHCKQLKSLSHIKK 480
Db 453 INKLSRLDKELDDYREESEYMAADEYNRLKQVKGSDYKSKNHCKQLKSLSHIKK 512

QY 481 MVGDYDRQKT 490
Db 513 MVGDYDRQKT 522

RESULT 4

US-10-028-780-8
; Sequence 8, Application US/10028780
; Publication No. US20020132329A1
; GENERAL INFORMATION:
; APPLICANT: DUAN, D. ROXANNE
; SHILATIPARD, ALI
; CONAWAY, JOAN W.
; TITLE OF INVENTION: ELL2, A New Member of an ELL Family of
; RNA Polymerase II Elongation Factors
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/028,780
; FILING DATE: 28-Dec-2001
; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/026,343
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 60/038,447
; FILING DATE: 19-FEB-1997

ATTORNEY/AGENT INFORMATION:
; NAME: GOLDSTEIN, JORGE A.
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 1488.0880001

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 372-2540

SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; STRANDEDNESS: No. US20020132329A1 Relevant
; TOPOLOGY: No. US20020132329A1 Relevant

MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-028-780-8

Query Match 21.9%; Score 566; DB 13; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.2e-38;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 383 DWIREYPPITSDQORQLYKRNFTGLQEQYKLSQSELDEINKLSRLDKELDDYRESSEY 442
Db 1 DWIREYPPITSDQORQLYKRNFTGLQEQYKLSQSELDEINKLSRLDKELDDYRESSEY 60

QY 443 MAAADEYNRLKQVKGSDYKSKNHCKQLKSLSHIKKMGVGDYDRQKT 490
Db 61 MAAADEYNRLKQVKGSDYKSKNHCKQLKSLSHIKKMGVGDYDRQKT 108

RESULT 5

US-10-262-839-158
; Sequence 158, Application US/10262839
; Publication No. US20040038877A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, John,
; APPLICANT: Anderson, David W.,
; APPLICANT: Boldog, Ferenc,
; APPLICANT: Burgess, Catherine,
; APPLICANT: Catterton, Elina,
; APPLICANT: Edinger, Shlomit,
; APPLICANT: Ellerman, Karen,
; APPLICANT: Gerlach, Valerie,
; APPLICANT: Gorman, Linda,
; APPLICANT: Guo, Xiaojia,
; APPLICANT: Ji, Weizhen,
; APPLICANT: Kekuda, Ramesh,
; APPLICANT: Leach, Martin,
; APPLICANT: Li, Li,
; APPLICANT: Miller, Charles,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Reiger, Daniel,
; APPLICANT: Rothenberg, Mark,
; APPLICANT: Shinkets, Richard,
; APPLICANT: Smithson, Glennda,
; APPLICANT: Spytek, Kimberly,
; APPLICANT: Taupier, Raymond, Jr.,
; APPLICANT: Vernet, Corine,
; APPLICANT: Voss, Edward,
; APPLICANT: Zerhusen, Brian,
; APPLICANT: Zhong, Mei

; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-462A
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/326,483
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/327,917
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,029
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,056
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/381,101
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/371,972
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/327,342
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 60/328,044
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,849
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/374,738
; PRIOR FILING DATE: 2002-04-23

; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 367
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 158
; LENGTH: 561
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-262-839-158

Query Match 11.2%; Score 291; DB 15; Length 561;
Best Local Similarity 22.0%; Pred. No. 5e-15;
Matches 110; Conservative 79; Mismatches 166; Indels 146; Gaps 18;

QY 12 YSPFPEDEILFYKWTSPPGVIRILSMILIVMCIAIFACVASTLAWDRGYGTSLLGGSYG 71
Db 172 YSEKVEEYNRLSYNKMWSWAGLLRILGVVLELLIGAGVACVTAITHKDSW-YNLFQYS-- 228


```
; APPLICANT: Symonds, James Matthew
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL
; FILE REFERENCE: 100086.407C7
; CURRENT APPLICATION NUMBER: US/10/006,869
; CURRENT FILING DATE: 2001-12-03
; NUMBER OF SEQ ID NOS: 4052
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Occludin Cell
; OTHER INFORMATION: Adhesion Recognition Sequence
US-10-006-869-62

Query Match 10.2%; Score 263; DB 14; Length 48;
Best Local Similarity 100.0%; Pred. No. 3.1e-14;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 GVNPTAQSSGSLGSIQYALCNQFYTPAATGLYVDQYLYHYCVDPQE 211
Db 1 GVNPTAQSSGSLGSIQYALCNQFYTPAATGLYVDQYLYHYCVDPQE 48

RESULT 10
US-10-193-653-69
; Sequence 69, Application US/10193653
; Publication No. US20030109454A1
; GENERAL INFORMATION:
; APPLICANT: Doherty, Patrick
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING ADHESION MOLECULE
; FILE REFERENCE: 100086.403C2
; CURRENT APPLICATION NUMBER: US/10/193,653
; CURRENT FILING DATE: 2002-07-10
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 69
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Solid Phase
US-10-193-653-69

Query Match 10.2%; Score 263; DB 14; Length 48;
Best Local Similarity 100.0%; Pred. No. 3.1e-14;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 GVNPTAQSSGSLGSIQYALCNQFYTPAATGLYVDQYLYHYCVDPQE 211
Db 1 GVNPTAQSSGSLGSIQYALCNQFYTPAATGLYVDQYLYHYCVDPQE 48

RESULT 11
US-10-395-032-62
; Sequence 62, Application US/10395032
; Publication No. US20030229199A1
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Symonds, James Matthew
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL
; FILE REFERENCE: 100086.407C9
; CURRENT APPLICATION NUMBER: US/10/395,032
; CURRENT FILING DATE: 2003-03-21
```

```
; NUMBER OF SEQ ID NOS: 4052
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Occludin Cell
; OTHER INFORMATION: Adhesion Recognition Sequence
US-10-395-032-62

Query Match 10.2%; Score 263; DB 15; Length 48;
Best Local Similarity 100.0%; Pred. No. 3.1e-14;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 GVNPTAQSSGSLGSIQYALCNQFYTPAATGLYVDQYLYHYCVDPQE 211
Db 1 GVNPTAQSSGSLGSIQYALCNQFYTPAATGLYVDQYLYHYCVDPQE 48

RESULT 12
US-10-425-557-78
; Sequence 78, Application US/10425557
; Publication No. US20040006011A1
; GENERAL INFORMATION:
; APPLICANT: Gour, Barbara J.
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Ali, Anmar
; APPLICANT: Ni, Feng
; APPLICANT: Chen, Zhigang
; APPLICANT: Michaud, Stephanie Denise
; APPLICANT: Wang, Shaomeng
; APPLICANT: Hu, Zengjian
; TITLE OF INVENTION: PEPTIDOMIMETIC MODULATORS OF CELL ADHESION
; FILE REFERENCE: 100086.401C16
; CURRENT APPLICATION NUMBER: US/10/425,557
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 78
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Occludin cell
; OTHER INFORMATION: adhesion recognition sequence and flanking amino
; OTHER INFORMATION: acids
US-10-425-557-78

Query Match 10.2%; Score 263; DB 15; Length 48;
Best Local Similarity 100.0%; Pred. No. 3.1e-14;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 GVNPTAQSSGSLGSIQYALCNQFYTPAATGLYVDQYLYHYCVDPQE 211
Db 1 GVNPTAQSSGSLGSIQYALCNQFYTPAATGLYVDQYLYHYCVDPQE 48

RESULT 13
US-10-412-701-78
; Sequence 78, Application US/10412701
; Publication No. US20040058864A1
; GENERAL INFORMATION:
; APPLICANT: Gour, Barbara J.
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Ali, Anmar
; APPLICANT: Ni, Feng
; APPLICANT: Chen, Zhigang
; APPLICANT: Michaud, Stephanie Denise
; APPLICANT: Wang, Shaomeng
; APPLICANT: Hu, Zengjian
; TITLE OF INVENTION: PEPTIDOMIMETIC MODULATORS OF CELL ADHESION
; FILE REFERENCE: 100086.413C2
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2005, 09:36:47 ; Search time 52.2335 Seconds
(without alignments) 902.605 Million cell updates/sec

Title: US-09-891-064A-2 COPY 33 522

Perfect score:

Sequence: 1 MHVRPMLSQPAYSFYPEDEI.....LKSKLSHIKKMVG DYDRQKT 490

Scoring table: BLOSUM62

Learning curve: 2000000
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

100% Precision: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79: *

1: pir1:*

```
2: pir2:*
```

3: dir3: *

4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	2590	100.0	522	2	G02533	occludin - human
2	1070.5	41.3	504	2	A49467	occludin - chicken
3	229.5	8.9	621	2	I38880	eleven-nineteen ly
4	142.5	5.5	591	1	KRWS2	keratin, type II c
5	140.5	5.4	643	1	KRHU2	keratin 1, type I
6	138.5	5.3	576	1	S22453	colicin E7 (SC 3.1
7	133	5.1	564	2	I61771	keratin 6f, type I
8	131	5.1	564	1	KRHUEA	keratin 6a, type I
9	129	5.0	564	1	KRHUEB	keratin 6b, type I
10	128.5	5.0	386	2	S51436	probable membrane
11	127	4.9	75	2	A89016	protein B0213.6 [i
12	126.5	4.9	564	2	I61770	keratin 6e, type I
13	126.5	4.9	944	2	S26710	spindle pole body
14	125.5	4.8	59	2	B99016	protein B0213.5 [i
15	125	4.8	553	2	I59009	epidermal keratin
16	123.5	4.8	180	2	JC7876	prion protein homo
17	123	4.7	1099	2	G90546	conserved hypothet
18	122.5	4.7	72	2	B89016	protein B0213.2 [i
19	122	4.7	278	1	J50638	osteopontin precur
20	121.5	4.7	226	2	T35435	probable integral
21	120.5	4.7	1699	2	T31340	voltage-gated sodi
22	119.5	4.6	564	2	I61768	keratin 6c, type I
23	119.5	4.6	1171	2	T45706	chromosome-associa
24	118.5	4.6	80	2	B45466	glycine/tyrosine-r
25	118	4.6	633	2	T41332	case homolog - fis
26	117.5	4.5	364	2	T31825	hypothetical prote
27	117	4.5	494	2	S10134	plasmid recombinat
28	116.5	4.5	2441	2	D71623	erythrocyte membra
29	115.5	4.5	486	1	KRXL	keratin 3, type I,

[illegible]

QY 322 LPLTSPVDDFRQPRYSGGFETPSKRAPAKGRAGRSKRTQDHYETDYYTTGGESCDELE 381
| | | : : : | | : : : | | : | | : | : |
Db 317 LVLDISIIEVK-----AQFEVIAQRSRAE-----AESLYQTYKEELQVTAGRHGDDL- 363
| | | : : : | | : : : | | : | | : | : |
QY 382 EDWIREYPPITSDQORQLYKRNFDTGLQFYKSLQSELDEINKELSRDLDDYRESEEE 441
| | | : : : | | : : : | | : | | : | : :
Db 364 -----RNTKQEIAEINRMIOQLRSEIDHVKKQCAN 393
| | | : : : | | : : : | | : | | : | : :
QY 442 YMAA-ADEYNR-----LKQVKGSA-----YKSKNHCKQLKS--KLSHIK 479
| | | : : : | | : : : | | : | | : | : :
Db 394 LQAATADAEQRGEMALKDARGKLEGLDALQAKAKODMAMLLKEYHELMNVK 444
| | | : : : | | : : : | | : | | : | : :

Search completed: April 1, 2005, 10:16:10
Job time : 55.2335 secs